JOURNAL OF CONSULTING PSYCHOLOGY

Edited By
Laurance F. Shaffer
Teachers College, Columbia University

Advisory Editors

Donald W. Fiske
University of Chicago

Harrison G. Gough
University of California

William A. Hunt
Northwestern University

E. Lowell Kelly
University of Michigan

Morris Krugman
Board of Education
New York, N. Y.

Bertha M. Luckey
Board of Education
Cleveland, Ohio

Fred McKinney
University of Missouri

Anne Roe
New York, N. Y.

Carl R. Rogers
University of Wisconsin

Nevitt Sanford
University of California

Edward Joseph Shoben, Jr.
Teachers College
Columbia University

George J. Wischner
University of Pittsburgh

J. R. Wittenborn
Rutgers University

VOLUME 22

1958

BUREAU OF EDNL. & PSYL. RESEARCH
LIBRARY.

Arthur C. Hoffman, *Managing Editor;* Helen Orr, *Promotion Manager;*
Barbara Cummings, *Editorial Assistant*

Published Bimonthly By

The American Psychological Association, Inc.
1333 Sixteenth St., N. W.
Washington 6, D. C.

Printed At
Lancaster, Pennsylvania

Copyright © 1958 by the American Psychological Association, Inc.

Contents of Volume 22

Aaronson, B. S. Age and Sex Influences on MMPI Profile Peak Distributions in an Abnormal Population 203

Altrocchi, J. See *Rosenberg, B. G.*

Altus, W. D. Q-L Variability, MMPI Responses, and College Males 367

Baughman, E. E., and Guskin, S. Sex Differences on the Rorschach 400

Becker, W. C. See *Frey, A. H.*

Beech, H. R., and Maxwell, A. E. Differentiation of Clinical Groups Using Canonical Variates 113

Bendig, A. W. Extraversion, Neuroticism, Radicalism, and Tendermindedness 292

Bendig, A. W. Extraversion, Neuroticism, and Verbal Ability Measures 464

Bendig, A. W. Identification of Item Factor Patterns Within the Manifest Anxiety Scale 158

Berg, I. A. See *Hesterly, S. O.*

Berger, D., Everson, R., Rutledge, L., and Koskoff, Y. D. The Spiral Aftereffect in a Neurological Setting 249

Block, J. An Unprofitable Application of the Semantic Differential 235

Bloomberg, R. See *Jackson, D. N.*

Bordin, E. S. See *Cutler, R. L.*

Brick, H. See *Lebo, D.*

Broen, W. E. Jr., and Wirt, R. D. Varieties of Response Sets 237

Brown, E. C., Casey, A., Fisch, R. I., and Neuringer, C. Trail Making Test as a Screening Device for the Detection of Brain Damage 469

Buss, A. H., and Durkee, Ann. Conditioning of Hostile Verbalizations in a Situation Resembling a Clinical Interview 415

Campbell, Mary M. See *Sumerwell, Harriet C.*

Carp, A. See *Himelstein, P.*

Cartwright, D. S. See *Kirtner, W. L.*

Casey, A. See *Brown, E. C.*

Chance, June Elizabeth. Adjustment and Prediction of Others' Behavior 191

Cohen, I. S. See *Corah, N. L.*

Corah, N. L., Feldman, M. J., Cohen, I. S., Gruen, W., Meadow, A., and Ringwall, E. A. Social Desirability as a Variable in the Edwards Personal Preference Schedule 70

Cowen, Judith E. See *Mandler, G.*

Cutler, R. L. Countertransference Effects in Psychotherapy 349

Cutler, R. L., Bordin, E. S., Williams, Joan, and Rigler, D. Psychoanalysts as Expert Observers of the Therapy Process 335

Daston, P. G. Stylus Maze Performance of Chronic Schizophrenics Taking Chlorpromazine 384

Davids, A. Intelligence in Childhood Schizophrenics, Other Emotionally Disturbed Children, and Their Mothers 159

Davis, L. See *Tuddenham, R. D.*

Davison, L. See *Tuddenham, R. D.*

deCharms, R. A Self-Scored Projective Measure of Achievement and Affiliation Motivation 172

Dennis, W. Handwriting Conventions as Determinants of Human Figure Drawings 293

Dilworth, Tom, IV. A Comparison of the Edwards PPS Variables with Some Aspects of the TAT 486

Dittmann, A. T. Problems of Reliability in Observing and Coding Social Interactions 430

Durkee, Ann. See *Buss, A. H.*

Eastman, D. Self Acceptance and Marital Happiness 95

Elkins, Elise. The Diagnostic Validity of the Ames "Danger Signals" 281

Eschenbach, A. E. See *Himelstein, P.*

Eskey, A., and Friedman, I. The Prognostic Significance of Certain Behavioral Variables - - 91
Everson, R. See *Berger, D.*
Feldman, M. J. See *Corah, N. L.*
Feldman, M. J., and Hunt, R. G. The Relation of Difficulty in Drawing to Ratings of Adjustment Based on Human Figure Drawings - - - 217
Fey, W. F. Doctrine and Experience: Their Influence Upon the Psychotherapist - - - - 403
Fiedler, F. E. See *Terwilliger, J. S.*
Fisch, R. I. See *Brown, E. C.*
Fisher, J. See *Little, K. B.*
Frey, A. H., and Becker, W. C. Some Personality Correlates of Subjects Who Fail to Appear for Experimental Appointments - - - 164
Friedman, I. See *Eskey, A.*
Geller, M. See *Gundlach, R. H.*
Gerstein, A. I. See *McKeever, W. F.*
Gertz, B. See *Stilson, D. W.*
Gilberstadt, H., Schein, J. D., and Rosen, A. Further Evaluation of the Archimedes Spiral Aftereffect - - - 243
Gisvold, D. A Validity Study of the Autonomy and Deference Subscales of the EPPS - - 445
Goldberg, L. R., and Smith, P. A. The Clinical Usefulness of the Archimedes Spiral in the Diagnosis of Organic Brain Damage - - - 153
Grams, A., and Rinder, L. Signs of Homosexuality in Human-Figure Drawings - - - - 394
Granick, S., and Scheflen, Norma A. Approaches to Reliability of Projective Tests with Special Reference to the Blacky Pictures Test - - - 137
Greenfield, N. S. Allergy and the Need for Recognition - - - 230
Greenfield, N. S. Personality Patterns of Patients Before and After Application for Psychotherapy - - - 280
Griffiin, Dorothy P. Movement Responses and Creativity - - - 134
Grigg, A. E. Experience of Clinicians, and Speech Characteristics and Statements of Clients as Variables in Clinical Judgment - - - 315
Grosz, H. J. See *Zuckerman, M.*
Gruen, W. See *Corah, N. L.*
Gundlach, R. H., and Geller, M. The Problem of Early Termination: Is it Really the Terminee? - - - 410
Gurrslin, O. See *Roach, J. L.*
Guskin, S. See *Baughman, E. E.*
Gynther, M. D. See *Stilson, D. C.*
Hamlin, R. M. See *Powers, W. T.*
Hannah, L. D. Causative Factors in the Production of Rotations on the Bender-Gestalt Designs - - - 398
Hare, A. P. See *Matarazzo, J. D.*
Hesterly, S. O., and Berg, I. A. Deviant Responses as Indicators of Immaturity and Schizophrenia - - - 389
Higgins, C., and Sivers, Cathryne H. A Comparison of Stanford-Binet and Colored Raven Progressive Matrices IQs for Children with Low Socioeconomic Status - - - 465
Hilden, A. H. Q-Sort Correlations: Stability and Random Choice of Statements - - - - 45
Hiler, E. W. An Analysis of Patient-Therapist Compatibility - - - 341
Himelstein, P., Eschenbach, A. E., and Carp, A. Interrelationships Among Three Measures of Need Achievement - - - 451
√*Howard, W.* A Note on McNemar's "On Abbreviated Wechsler-Bellevue Scales" - - - - 414
Hunt, D. E., and Schroder, H. M. Assimilation, Failure-Avoidance, and Anxiety - - - - 39
Hunt, R. G. See *Feldman, M. J.*
Hunt, R. G. See *Roach, J. L.*

Hyman, M., and Wohl, J. Environmental Factors and Outpatient Clinic Intake - 431
Jackson, D. N., and Bloomberg, R. Anxiety: Unitas or Multiplex? - 225
Johnson, L. C. Rorschach Concept Evaluation Test as a Diagnostic Tool - 129
Jones, M. B. The Polarity of Psychological Tests - 25
Kantor, R. E. See *Winder, C. L.*
Karson, S., and Pool, K. B. Second-Order Factors in Personality Measurement - 299 ✓
Kasper, S. Progressive Matrics (1938) and Emotional Disturbance - 24
Katz, L. Learning in Aphasic Patients - 143
Katz, M. M. See *Lorr, M.*
Katz, M. M., Lorr, M., and Rubinstein, E. A. Remainer Patient Attributes and Their Relation to Subsequent Improvement in Psychotherapy - 411
Kausler, D. H. See *Trapp, E. P.*
Kausler, D. H., and Trapp, E. P. Relationship Between Achievement Motivation Scores and Manifest Anxiety Scores - 448
Kelleher, D. The Social Desirability Factor in Edwards' PPS - 100
Kimble, Gloria. See *Spilka, B.*
King, F. W. A Normative Note on Sentence Completion Cross-Sex Identification Responses - 63
Kirtner, W. L., and Cartwright, D. S. Success and Failure in Client-Centered Therapy as a Function of Client Personality Variables - 259
Kirtner, W. L., and Cartwright, D. S. Success and Failure in Client-Centered Therapy as a Function of Initial In-Therapy Behavior - 329
Koskoff, Y. D. See *Berger, D.*
Krasner, L. A Technique for Investigating the Relationship Between the Behavior Cues of the Examiner and the Verbal Behavior of the Patient - 364
Krugman, A. D. See *Vernier, Claire.*
Lauterbach, C. G. The Taylor A Scale and Clinical Measures of Anxiety - 314
Lebo, D. A Theoretical Framework for Nondirective Play Therapy - 275
Lebo, D., Toal, R. A., and Brick, H. Manifest Anxiety in Prisoners Before and After CO_2 - 51
Levine, M. "Not Alike" Responses in Wechsler's Similarities Subtest - 480
Linde, T., and Patterson, C. H. The MMPI in Cerebral Palsy - 210
Little, K. B., and Fisher, J. Two New Experimental Scales of the MMPI - 305
Lorr, M. See *Katz, M. M.*
Lorr, M., Katz, M. M., and Rubinstein, E. A. The Prediction of Length of Stay in Psychotherapy - 321
McKeever, W. F., and Gerstein, A. I. Validity of the Hewson Ratios: Investigation of a Fundamental Methodological Consideration - 150
Mandler, G., and Cowen, Judith E. Text Anxiety Questionnaires - 228
Manis, M. Personal Adjustment, Assumed Similarity to Parents, and Inferred Parental-Evaluations of the Self - 481
Marcuse, F. L. See *Martin, R. M.*
Mark, H. J., and Pasamanick, B. Asynchronism and Apparent Movement Thresholds in Brain-Injured Children - 173
Marlowe, D. Some Psychological Correlates of Field Independence - 334
Martin, R. M., and Marcuse, F. L. Characteristics of Volunteers and Nonvolunteers in Psychological Experimentation - 475
Mason, D. J. See *Stilson, D. W.*
Matarazzo, J. D., Saslow, G., and Hare, A. P. Factor Analysis of Interview Interaction Behavior - 419
Maxwell, A. E. See *Beech, H. R.*
Meadow, A. See *Corah, N. L.*
Mednick, S. A. See *Wertheim, J.*

Melikian, L. H. The Relationship Between Edwards' and McClelland's Measures of Achievement Motivation 296
Miller, Christine. See *Sullivan, P. L.*
Mitsos, S. B. Representative Elements in Role Construct Technique 311
Monashkin, I. See *Zuckerman, M.*
Motto, J. J. The MMPI Performance of Veterans with Organic and Psychiatric Disabilities 304
Murstein, B. I. Nonprojective Determinants of Perception on the TAT 195
Murstein, B. I. The Relationship of Stimulus Ambiguity on the TAT to the Productivity of Themes 348
Murstein, B. I. Some Determinants of the Perception of Hostility 65
Nelson, P. See *Willingham, W. W.*
Neuringer, C. See *Brown, E. C.*
Nicholas, Alma L. See *Reznikoff, M.*
Norton, J. A., Jr. See *Zuckerman, M.*
Nunnally, J. Opinions of Psychologists and Psychiatrists About Mental Health Problems 178
O'Connor, W. See *Willingham, W. W.*
Oltean, Mary. See *Zuckerman, M.*
Pasamanick, B. See *Mark, H. J.*
Patterson, C. H. See *Linde, T.*
Pool, K. B. See *Karson, S.*
Powers, W. T., and Hamlin, R. M. A Comparative Analysis of Deviant Rorschach Response Characteristics 123
Psychological Test Reviews 82, 241, 497
Rapaport, G. M. "Ideal Self" Instructions, MMPI Profile Changes, and the Prediction of Clinical Improvement 459
Rehfisch, J. M. A Scale for Personality Rigidity 11
Rehfisch, J. M. Some Scale and Test Correlates of a Personality Rigidity Scale 372
Rempel, P. P. The Use of Multivariate Statistical Analysis of Minnesota Multiphasic Personality Inventory Scores in the Classification of Delinquent and Nondelinquent High School Boys 17
Reznikoff, M., and Nicholas, Alma L. An Evaluation of Human-Figure Drawing Indicators of Paranoid Pathology 395
Reznikoff, M., and Toomey, Laura C. The Weighted Q Sort. A Procedure for Quantitatively Estimating Emotional Disturbance and Personality Change 187
Ribback, Beatrice B. See *Zuckerman, M.*
Rigler, D. See *Cutler, R. L.*
Rinder, L. See *Grams, A.*
Ringwall, E. A. See *Corah, N. L.*
Roach, J. L., Gurrslin, O., and Hunt, R. G. Some Social-Psychological Characteristics of a Child Guidance Clinic Caseload 183
Roberts, A. H. Chance Frequency in Matching Problems When Success or Failure is Reported After Each Matching Operation 233
Rosen, A. Differentiation of Diagnostic Groups by Individual MMPI Scales 453
Rosen, A. See *Gilberstadt, H.*
Rosenberg, B. G., and Altrocchi, J. The Yacorzynski Block Technique: A Cross-Validation Study 122
Ross, A. O. Brain Injury and Intellectual Performance 151
Rubinstein, E. A. Analysis of Self and Peer Personality Ratings of Psychotherapists and Comparison with Patient Ratings 10
Rubinstein, E. A. See *Katz, M. M.*
Rubinstein, E. A. See *Lorr, M.*
Rutledge, L. See *Berger, D.*

Salcines, R. A. See *Winter, W. D.*
Sarason, I. G. See *Sumerwell, Harriet C.*
Saslow, G. See *Matarazzo, J. D.*
Saucer, R. T. A Further Study of the Perception of Apparent Motion by Schizophrenics - 256
Scheflen, Norma A. See *Granick, S.*
Schein, J. D. See *Gilberstadt, H.*
Scherer, I. W. See *Trehub, A.*
Schindler, R. See *Tuddenham, R. D.*
Schroder, H. M. See *Hunt, D. E.*
Scott, E. M. A Case of Folie a Deux and Projective Techniques - - - - - - - - - - 90
Sharp, H. C. A Note on the Reliability of the Leiter International Performance Scale - - 320
Siegman, A. W. The Effect of Cultural Factors on the Relationship Between Personality, Intelligence, and Ethnocentric Attitudes - - - - - - - - - - - - - - - - - - 375
Sivers, Cathryne H. See *Higgins, C.*
Smelser, W. See *Sullivan, P. L.*
Smith, G. M. Six Measures of Self-Concept Discrepancy and Instability: Their Interrelations, Reliability, and Relations to Other Personality Measures - - - - - - - - - 101
Smith, P. A. See *Goldberg, L. R.*
Sommer, Dorothy T. See *Sommer, R.*
Sommer, R., and Sommer, Dorothy T. Assaultiveness and Two Types of Rorschach Color Responses - 57
Sopchak, A. L. Spearman Correlations Between MMPI Scores of College Students and Their Parents - 207
Sperrazzo, G., and Wilkins, W. L. Further Normative Data on the Progressive Matrices - 35
Spielberg, Mimi J. See *Tucker, J. E.*
Spielberger, C. D. On the Relationship Between Manifest Anxiety and Intelligence - - - 220
Spilka, B., and Kimble, Gloria. Personality Correlates of Q-L Differentials on the ACE - - 142
Stafford, J. W. See *Vernier, C. M.*
Stennett, R. G., and Thurlow, M. Cultural Symbolism: The Age Variable - - - - - - 496
Stilson, D. W., Mason, D. J., Gynther, M. D., and Gertz, B. An Evaluation of the Comparability and Reliabilities of Two Behavior Rating Scales for Mental Patients - - - 213
Strupp, H. H. The Performance of Psychoanalytic and Client-Centered Therapists in an Initial Interview - 265
Sullivan, P. L., Miller, Christine, and Smelser, W. Factors in Length of Stay and Progress in Psychotherapy - 1
Sumerwell, Harriet C., Campbell, Mary M., and Sarason, I. G. The Effect of Differential Motivating Instructions on the Emotional Tone and Outcome of TAT Stories - - - - 385
Tamkin, A. S. Rorschach Card Rejection by Psychiatric Patients - - - - - - - - - 441
Tatom, Mary Helen. A Factorial Isolation of Psychiatric Outpatient Syndromes - - - - 73
Taulbee, E. S. Relationship Between Certain Personality Variables and Continuation in Psychotherapy - 83
Terwilliger, J. S., and Fiedler, F. E. An Investigation of Determinants Inducing Individuals to Seek Personal Counseling - 288
Thune, L. E. See *Whitehead, W. A.*
Thurlow, M. See *Stennett, R. G.*
Toal, R. A. See *Lebo, D.*
Toomey, Laura C. See *Reznikoff, M.*
Trapp, E. P. See *Kausler, D. H.*
Trapp, E. P., and Kausler, D. H. Text Anxiety Level and Goal-Setting Behavior - - - 31
Trehub, A., and Scherer, I. W. Wechsler-Bellevue Scatter as an Index to Schizophrenia - - - 147
Trier, T. R. Vocabulary as a Basis for Estimating Intelligence from the Rorschach - - - 289

Tucker, J. E., and Spielberg, Mimi J. Bender-Gestalt Test Correlates of Emotional Depression 56
Tuddenham, R. D., Davis, L., Davison, L., and Schindler, R. An Experimental Group Version for School Children of the Progressive Matrices 30
Vernier, Claire M., Stafford, J. W., and Krugman, A. D. A Factor Analysis of Indices from Four Projective Techniques Associated with Four Different Types of Physical Pathology 433
Wahler, H. J. Social Desirability and Self-Ratings of Intakes, Patients in Treatment, and Controls 357
Walters, R. H., and Zaks, M. S. Changes in Responses to a Set of Personality-Inventory Items as a Function of Age 458
Wattron, J. B. Validity of Marsh-Hilliard-Liechti MMPI Sexual Deviation Scale in a State Prison Population 16
Wertheim, J., and Mednick, S. A. The Achievement Motive and Field Independence 38
Whitehead, W. A., and Thune, L. E. The Effects of Chlorpromazine on Learning in Chronic Psychotics 379
Wilkins, W. L. See *Sperrazzo, G.*
Williams, Joan. See *Cutler, R. L.*
Williams, W. C. The PALS Tests: A Technique for Children to Evaluate Both Parents 487
Willingham, W. W., Nelson, Paul, and O'Connor, W. A Note on the Behavioral Validity of the Gordon Personal Profile 378
Winder, C. L., and Kantor, R. E. Rorschach Maturity Scores of the Mothers of Schizophrenics 438
Winter, W. D., and Salcines, R. A. The Validity of the Objective Rorschach and the MMPI 199
Wirt, R. D. See *Broen, W. E., Jr.*
Wohl, J. See *Hyman, M.*
Zaks, M. S. See *Walters, R. H.*
Zuckerman, M., and Grosz, H. J. Suggestibility and Dependency 328
Zuckerman, M., Oltean, Mary, and Monashkin, I. The Parental Attitudes of Mothers of Schizophrenics 307
Zuckerman, M., Ribback, Beatrice B., Monashkin, I., and Norton, J. A., Jr. Normative Data and Factor Analysis on the Parental Attitude Research Instrument 165

Journal of Consulting Psychology
Vol. 22, No. 1, 1958

Factors in Length of Stay and Progress in Psychotherapy[1]

Patrick L. Sullivan,
Edward Glaser & Associates, San Francisco

Christine Miller,
VA Mental Hygiene Clinic, Oakland

and William Smelser
VA Hospital, Oakland

Research in the area of prediction of outcome of psychotherapy has produced mainly contradictory findings and at the best only low degrees of predictive accuracy for any one variable. The reason probably lies in the fact that a wealth of factors may affect outcome in therapy. For the sake of convenience, these may be roughly classified into three groups: (*a*) characteristics of the patient; (*b*) characteristics of the therapist; (*c*) situational variables, such as clinic policies as to length of treatment, eligibility requirements, and external factors in the patient's life.

Because so many factors may be important in determining outcome of therapy in each one of these three groups, it seems unreasonable to expect more than a low degre of predictability for any one factor which may be selected for study. Also, when one cross-validates the findings of a previous investigator or changes the setting, factors of relevance to the outcome of therapy are almost certain to change in each of the three categories. Thus we find one investigator, Gallagher (5), who reported that high scores on Depression on the MMPI are associated with "success" in psychotherapy, while Barron (2) found that high Depression scores are associated with lack of improvement in psychotherapy. The possibilities for variation in factors of relevance to outcome of therapy are so numerous that the difficulties in generalizing from one setting to another seem almost insurmountable.

These thoughts have been generated from our experience in a study of factors related to length of stay and progress in treatment. After having several hypotheses confirmed in one sample, we undertook to cross-validate our findings with two other samples in the same setting. The inconsistency of most of our findings from one sample to another is of perhaps as much interest as a few of our findings which were consistent.

The purposes of this study were threefold:

1. We wished to evaluate the relationship of social status variables to length of stay and progress in psychotherapy in a VA mental hygiene clinic. Previous investigators (1, 8, 9, 10) had found that high social status was associated with favorable outcome in therapy, and we wished to ascertain the extent to which this relationship would hold in our clinic.

2. We wished to explore some of the psychological variables which might be associated with social status in the hope of gaining a better understanding of what makes social status a predictor of outcome in psychotherapy. Special scales from the MMPI were used in this endeavor, and these will be described below. We were also interested in any psychological differences between groups having favorable and unfavorable outcomes in therapy which would come to light on the MMPI.

---

[1] Approved for publication by Veterans Administration Regional Office, San Francisco.

3. We wished to examine available data on therapists in relation to length of stay and improvement.

Method

The data for the present study were gathered from the records of patients in the Veterans Administration Mental Hygiene Clinic, Oakland, California. Only the records of male, white veterans of either World War II or the Korean action were included. Data from patients who terminated treatment by reason of legal ineligibility were not included. Our total sample consisted of 268 patients, of whom 196 had completed the Minnesota Multiphasic Personality Inventory at the time of initial contact with the clinic.

Data concerning the psychodiagnostic categories of the patients are not presented. The diagnoses were examined by the investigators, and it was concluded that they were too unreliable and too contaminated with legalistic procedures to provide a sound basis for grouping. It may be noted, however, that only approximately 25% of the patients were designated as schizophrenic, so that our findings apply predominantly to psychoneurotic and psychosomatic cases.

The study was begun by gathering the records of cases closed in 1952–53, which numbered 131 (Group A). These data were analyzed and the findings checked against a sample drawn in a similar manner from the 1951 closings. It became apparent that an insufficient number of cases was available for the purpose (N is 43 for this second group, hereafter designated Group B), so an additional sample was drawn from 1954–55 closings. This last sample consisted of 94 cases and will be called Group C.

The variables chosen for investigation may be divided into four groups as follows:

A. Demographic measures:

a. Age.

b. Education. The number of years of schooling completed was translated into a seven-point scale in accordance with Warner's Code (11). In this code high school graduates are coded three, college undergraduates two, and college graduates one.

c. Occupation. Vocations were also transformed into scale scores following the above scheme, and again the relationship is an inverse one.

d. Education plus Occupation. The scaled scores of the two single measures were added together to approach a more comprehensive measure of socioeconomic status.

e. Education minus Occupation. The difference in coded values for Education and Occupation provided a measure of the discrepancy between education and vocation.

B. Psychological test measures ($N = 196$):

a. MMPI clinical and validating scales.

b. Special MMPI scales:

1. Social Status (developed by Gough [6] and found by him to be an accurate reflection of actual social status).

2. Intellectual Efficiency (developed by Gough [7]. This was used as our only available measure of intellectual functioning. Scores on this scale have been shown to be significantly related to high school and college grades).

3. Ego Strength (derived by Barron [3] by selection of items which differentiated a group of patients who improved in psychotherapy from those who did not improve).

4. Factor A (a measure of anxiety and general maladjustment developed by Welsh [12]).

5. Factor R (a measure of both the ability and the tendency to repress developed by Welsh [12]).

C. Therapist measures:

a. Sex.

b. Professional discipline (psychiatrist, psychologist, social worker).

c. Experience. An assignment of the therapists was made to "Experienced" if there was at least one year of staff work; otherwise, they were designated "Inexperienced."

D. Outcome measures:

a. Length of stay. The number of individual therapy interviews was used. In the vast majority of cases, interviews were held weekly, and that frequency may be taken as the definition of intensity of treatment for this study. The cases were split at the median number of interviews and designated Stay and Not-stay.

b. Improvement ratings. Ratings on a five-point scale were made by the individual therapists after termination. One point indicated "worse," and the other four points ranged from "no change" (two) to "much improved" (five). Only the cases which were above the median number of interviews, i.e., the Stay group, were included in the Improved and Not-improved data analysis. Patients who received the lowest three points on the rating scale were categorized as "Unimproved" and the others as "Improved." It may be noted that these ratings, depending as they do wholly upon the judgment of the treating therapist, are of unknown reliability and validity. Generalizations from the results must be regarded as exploratory only. In a sense, they may be looked upon as a measure of therapists' habits in rating improvement. The number of interviews, on the other hand, is a more objective measure.

Results

Length of Stay

Table 1 contains the results for the 131 cases constituting Group A, divided into Stay and Not-stay subgroups by dichotomizing at nine interviews, the median. Many measures significantly differentiate the two subgroups, with Education emerging as the most effective single variable. The results of both the demographic and psychological measures show that the patients who were better integrated and more successful in life pursuits were the ones who remained in treatment longer. The Stay group is higher on Education and Occupation. The MMPI findings show the Stayers to be higher on *IE,* Social Status, Ego Strength, and on *R* and *K*. Both of the latter may be regarded as reflecting the ability to hold together defenses effectively. *Pa,* on the other hand, is higher in the Not-stay group, which, along with the *A* measure, indicates greater maladjustment in that group. In the main, a general kind of psychological health appeared to be associated with the higher social status characteristics of the Stay group.

In Group A, most of our expectations with regard to psychological differences on the MMPI between Stayers and Not-stayers were confirmed. In contrast with Group A, the findings on psychological variables in Groups B and C shift markedly and are not consistent with those of Group A.

In Table 2 are given the means, *CR*'s and *t*'s where significant, for Groups B and C. Nonsignificant differences are not shown. The median number of interviews for Group B was identical with that obtained for Group A, i.e., nine, but was only five for Group C. No cause for this drop has been discovered, for there were no changes in the clinic's treatment policy nor any important alterations in the eligibility requirements.

Since the inconsistencies between groups might have followed from their being dichoto-

Table 1

Mean Differences Between Group A Stay and Non-stay Subgroups

	Stay		Non-stay			
Variable	Mean	*SD*	Mean	*SD*	*CR*	*p*
Demographic						
Age	30.7	5.9	32.9	6.9	2.0	.05
Educational rating (Ed)	2.92	1.4	3.72	1.1	3.68	.001
Occupational rating (Occ)	4.33	1.4	4.82	1.2	2.23	.02
Ed plus Occ	7.23	2.3	8.55	2.1	3.15	.001
Ed minus Occ	−1.41	1.4	−1.06	1.3	1.45	.10
MMPI special scales						
Social status	20.1	4.1	19.0	4.2	1.31	.10
Ego strength	41.5	12.5	38.5	7.3	1.47	.07
Intellectual efficiency	40.3	7.1	36.6	8.2	2.39	.01
A	19.1	7.8	22.3	8.4	1.75	.04
R	18.3	4.0	16.5	5.8	1.80	.04
MMPI regular scales						
L	47.9	8.1	48.1	4.8		ns
F	59.7	10.2	60.9	10.9		ns
K	52.4	10.3	48.1	8.9	2.56	.01
Hs	76.6	15.8	75.7	15.2		ns
D	79.5	15.4	82.1	18.1		ns
Hy	72.7	8.9	73.0	13.0		ns
Pd	67.0	11.5	66.4	14.1		ns
Mf	62.8	14.1	60.4	16.0		ns
Pa	56.7	10.6	61.7	13.5	2.35	.02
Pt	72.1	15.5	73.3	15.7		ns
Sc	68.2	16.1	69.5	17.1		ns
Ma	55.0	9.2	54.0	12.0		ns

Table 2

Cross-Validation of Differences Between Stay and Non-stay Subgroups in Groups B and C

Group and Variable	Stay		Non-stay			
	Mean	*SD*	Mean	*SD*	*t*	*p*
Group B						
Education	2.85	.93	3.7	1.12	2.65	.02
Occupation	4.24	1.74	4.67	1.36		ns
Ed plus Occ	7.08	2.62	8.38	1.90	1.81	.10
Ed minus Occ	−1.39	1.18	−.97	1.34		ns
Pt	78.8	[a]	70.7	[a]	2.1	.05
Sc	75.4	[a]	66.3	[a]	1.88	.10
Group C						
Ego strength	36.8	10.3	41.0	6.2	2.40	.02
Hs	79.3	18.8	71.3	11.5	2.57	.02
D	85.2	15.8	76.4	18.3	2.56	.02

[a] *SD*s not given because data, unfortunately, were discarded after *t*'s had been computed.

mized at different points, it was decided, therefore, to re-examine the data for Group C, making the division into subgroups not at five but nine interviews. The results of this revised alignment of cases are given in Table 3. The greatest changes that result from this shift are in the means of the demographic variables. These results are much more similar to those of Group A, but none of them reaches significance. Among the MMPI scales, only minor differences in means between the original and revised subgroups of Group C are effected.

Because none of our samples was very large, the data for Groups A, B, and C were combined to give a larger *N*, using the revised assignment of cases for Group C. The results are shown in Table 4. There are no significant differences between the Stay and Not-stay groups on the MMPI scales; in fact, the profiles of the Stay and Not-stay group are strikingly similar. This finding was to be expected because of the few significant differences within the individual samples and the frequent reversals of direction. Differences in demographic measures, on the other hand, are all significant except Education minus Occupation, Education being the most differentiating among them.

Table 3

Differences Between Stay and Non-stay in Group C
(Revised alignment)

Demographic Variables	Stay		Non-stay	
	Mean	*SD*	Mean	*SD*
Age	33.3	8.06	34.1	7.68
Educational rating	2.86	1.15	3.09	1.27
Occupational rating	4.30	2.47	4.97	1.84
Ed plus Occ	7.36	3.05	8.05	2.80
Ed minus Occ	−1.46	1.40	−1.87	1.46

Note.—All differences are nonsignificant.

While we confirmed the findings of other investigators that social status variables are relevant to length of stay, we failed to identify any psychological variables which consistently differentiate the Stayers from Not-stayers.

Therapeutic Improvement

It will be recalled that the analysis of data for factors related to progress in treatment was limited to those cases included in the Stay category. Also, the division into Improved and Unimproved was effected by subsuming under the former those cases rated in the highest two of the five points on the rating scale.

The handling of the data paralleled that described for the length-of-stay material. Because the *N*s in groups A, B, and C taken separately are so small, only the combined data for an *N* of 98 for demographic variables and 83 for the MMPI variables are presented. It should be noted, however, that compared with the frequent reversals from year to year for the Stay and Not-stay groups, there was much more consistency from sample to sample on the MMPI variables. Table 5 presents the findings for the total group of Improved compared with Unimproved. In marked contrast to the finding that the Stay and Not-stay groups were almost identical in MMPI profiles, nearly all the clinical scales show significant differences between the Improvers and Unimprovers. The Improvers have generally lower profiles than the Unimprovers, suggesting as Barron has said that those who improve in psychotherapy are less sick in the first place (2).

Turning to the demographic variables, the means of the Improved patients are seen to have contributed disproportionately to the differences found between the Stay and Not-stay groups. That is, the Improved show the Stay tendencies to a greater degree than do the Unimproved. The mean of Education for the Stay group contained in Table 4 is 2.84, for example, while the Improved group mean in Table 5 is 2.67; the Improved are better educated than the Stayers as a whole.

As the psychological findings indicated that the Improved were less disturbed at the outset of therapy than the Unimproved, so do the findings on the demographic variables show that they also held better positions and were employed in capacities more closely in keeping with their higher educational achievement.

Turning to other findings not covered in the foregoing tables, it might first be noted that there is a highly significant difference between the means of the number of interviews of the Improved and Unimproved. The means are 31.9 and 21.3, respectively, which, with standard deviations of 24.1 and 15.9, yield a *CR* of 2.5 ($p < .01$). In view of the manner in which our progress ratings were

Table 4

Total Group (A, B, and C), Differences Between Stay and Non-stay

Variable	Stay Mean	Stay *SD*	Non-stay Mean	Non-stay *SD*	*CR*	*p*
Demographic						
Age	31.4	5.8	33.2	7.48	2.19	.05
Educational rating (Ed)	2.84	1.19	3.42	1.26	3.74	.001
Occupational rating (Occ)	4.27	1.87	4.90	1.82	2.74	.01
Ed plus Occ	7.20	2.69	8.30	2.43	3.52	.001
Ed minus Occ	−1.43	1.35	−1.46	1.48		ns
MMPI special scales						
Social status	20.0		19.4			ns
Ego strength	39.0		39.1			ns
Intellectual efficiency	39.0		38.4			ns
A	20.1		20.6			ns
R	18.1		17.0			ns
MMPI regular scales						
L	47.8		47.7			ns
F	61.3		60.8			ns
K	51.2		49.1			ns
Hs	76.3		74.1			ns
D	82.0		80.5			ns
Hy	73.8		73.0			ns
Pd	67.5		68.2			ns
Mf	63.7		62.5			ns
Pa	58.9		61.7			ns
Pt	73.9		72.2			ns
Sc	71.6		69.6			ns
Ma	56.9		57.3			ns

Table 5

Total Group (A, B, and C), Differences Between Improved and Unimproved Cases

Variable	Improved Mean	Improved *SD*	Unimproved Mean	Unimproved *SD*	*CR*	*p*
Demographic						
Age	31.0	5.8	31.5	6.4		ns
Educational rating	2.67	1.31	2.97	1.36		ns
Occupational rating	3.89	1.33	4.65	1.30	2.84	.01
Ed plus Occ	6.56	1.70	7.60	2.46	1.97	.10
Ed minus Occ	−1.18	0.98	−1.68	.94	2.57	.01
MMPI special scales						
Social status	20.2		19.5			ns
Ego strength	41.8		38.5			ns
Intellectual efficiency	40.1		38.4			ns
A	18.0	7.3	21.8	6.6	2.69	.01
R	18.2		17.9			ns
MMPI regular scales						
L	49.0		46.8			ns
F	57.0	8.4	64.8	7.5	4.82	.0001
K	52.8		50.1			ns
Hs	74.1	7.5	78.1	8.3	2.5	.02
D	77.2	11.5	85.7	12.2	3.54	.001
Hy	70.4	5.1	76.4	5.4	5.65	.0001
Pd	64.8	9.3	69.3	8.6	2.46	.02
Mf	63.3		63.9			ns
Pa	56.7		60.6			ns
Pt	70.4	10.7	76.9	11.8	2.84	.01
Sc	67.4	11.4	75.3	13.6	3.11	.001
Ma	54.7	7.1	58.6	7.1	2.70	.01

obtained—i.e., from the treating therapist only—the determination of whether this finding is a "therapist" or "patient" phenomenon can only be left to speculation.

The usefulness of single measures for individual prediction was closely examined. None of them provide an adequate basis for such a purpose, although Education approached significance for length-of-stay predictions. If the large group of high school graduates is eliminated, the groups with less than high school versus those with more, yield highly significant differences with regard to length of stay.

Therapist Variables

It was indicated earlier that measures of therapist variables were also studied. None of the results of this analysis of therapist variables was significant. Therapists' sex and professional training showed only chance relationships with length-of-stay and progress ratings. Experienced therapists showed a systematic but nonsignificant inclination to (*a*) keep patients in treatment longer and (*b*) rate marked improvement more frequently.

It was also found, though again not consistently through the various samples, that more experienced therapists were assigned patients with better prognosis, i.e., with higher educational and occupational standing. In order to learn whether the superior outcomes achieved by experienced therapists was a function of experience or of better patient prognosis, the following analysis was made: Patients were divided into High and Low Education groups. Chi squares were computed by comparing, first, Stay, Not-stay proportions between Experienced and Inexperienced therapists within the High Education group; then the same comparison was made for the Low Education group. A similar analysis was made for the Improved and Unimproved groupings. It was found that the Experienced Therapist–High Education combination showed a higher probability for better outcome. The results were, however, far short of statistical significance. (*p* equals approximately .30 for both length-of-stay and progress.) The Low Education group had equally good or bad outcome regardless of therapist experience.

Discussion

The failure of the findings concerning length of stay of Group A to be substantiated in the other samples suggests several lines of thought. The most obvious implication to be drawn is that findings in this research area must be viewed with considerable tentativeness until thoroughly validated. This is particularly true of psychological test findings, as Windle has pointed out in his excellent review of the literature (13).

We might also take cognizance of other ways in which this study fails to meet standards of a comprehensive and rigorous investigation. The lack of objectivity in the therapeutic outcome ratings has already been noted, as has the fact that adequate diagnoses were not available. The amount of diagnostic homogeneity of our patient population can only be speculated on, and the differences in our measures among diagnostic categories is unknown. It has not been possible to describe the variations in competence and approach among therapists, nor is there information about such important items as date of onset of initial symptoms. Finally, our ratings do not encompass a follow-up period at all.

Despite these limitations, our results corroborate previous studies and add further substantiation to a picture which is evolving with ever-increasing clarity. The core of this is that those persons who are least equipped to meet life challenges are the ones who stand to gain least from psychotherapy.

Aside from considerations of the situation and how it might be selectively contributing to the results, our findings highlight the difference between "ability to use" and "need for" psychotherapy. These are not new concepts. They are obviously implicit in the criteria operating, for example, in the selection of patients for psychoanalysis during the trial analytic phase. In other treatment contexts, however, such as the one in which this study was conducted, the importance of the selection of "good" patients is sometimes obscured. This has its commendable aspects, in that many patients who might be turned away from other treatment facilities are accepted and efforts made to rehabilitate them. It is the personal experience of the writers that such efforts expand considerably the horizons of the professional worker and force the evolution of thinking that might never be produced were the selection of patients more rigorous. There are, furthermore, salutary effects achieved in a minority of cases. The worth of maintaining persons on jobs, in their own homes, and preventing them from becoming wards of hospitals cannot be gainsaid.

The fact remains that these two concepts are inversely related in our population, and alleviation is not being effected in many of those cases where it is most pressingly needed. Quite contrary to the requirements and, implicity at least, the expectations of society, the patients most in "need" are being helped minimally if at all.

If these persons are to be extended aid which may realistically be expected to be of benefit to them, it would appear that quite different approaches to their treatment will have to be evolved. In the development of these new attacks on the problem, the principal finding of this and other research (8, 9, 10) provide adequate starting points. In particular, if psychotherapy is characterizable as a process which places great importance on the ability of the patient to function at a high level of abstraction, the ability of the potential patient to engage in this process will have to be more precisely determined. The finding with respect to Education supports the inference that those with more schooling are both more intelligent and more capable of making abstractions about themselves. The research of Eells *et al.* (4) into the relatively lower abstractional ability of lower socioeconomic groups is pertinent to ours and to the findings of the Yale group (9). An approach such as the foregoing would be aimed at identifying those patients who are capable of entering into treatment as it is presently conducted by the majority of professional persons.

What are the implications of these findings for patients who do not demonstrate the ability to function at a high level of abstraction? The authors do not pretend to have comprehensive answers to this question but can at least extend suggestions. For one thing, it would be advisable to structure treatment at the outset for these patients in such a way as to concretize it. Specific time limitations on the length of treatment could be discussed at the time of intake and definite problem areas would be mutually agreed on and adhered to. Periodic summing-up, focused on the problem areas initially delineated, might be introduced. The effort to rehabilitate would be steered by the therapist with minimal demands on self-awareness and would focus instead on more concrete alternatives to the patients' current perceptions and behaviors. In short, therapy would be tailored to suit the individual's functioning ability to deal with gradated levels of abstraction.

Our study also illustrates the difficulties encountered in this area of research. Cross-validations in the same setting failed to confirm most of the original findings and pointed up the difficulties of generalizing results even from one year to the next. After considering our own difficulties in getting stable findings, we propose a few suggestions for further research in this area, most of which seem obvious, but which nevertheless have not often been adhered to.

1. Cross-validation of findings is essential in this area of research. An unexpectedly large proportion of the obtained "significant" differences may vanish on repetition with successive patient groups in the same setting. It seems wise, therefore, to regard findings until they have been cross-validated as only interesting, possibly chance findings which need to be further tested.

2. Large samples seem imperative because of the impossibility of controlling or even specifying many of the innumerable factors which can affect outcome in therapy.

3. As many as possible of the conditions of the study—the characteristics of patients and therapists—should be specified to aid in the evaluation of findings.

4. The study of groups which are homogeneous with regard to factors related to prognosis will enable meaningful differences to come to light in the context in which they are relevant.

5. It may be that we will come to a viewpoint prevalent in industrial psychology that local norms are necessary for sufficiently accurate prediction. Preliminary work may help to delineate some broad factors of relevance to outcome in therapy. It may be that from there on our efforts will be best expended in developing norms for our particular settings, patient populations, and possibly therapists.

Conclusion

A study of the factors related to length of stay and progress in psychotherapy in a VA mental hygiene clinic was carried out. Particular attention was given to measures of social status and to psychological variables which might be associated with social status. It was found that while the Stay and Not-stay groups did not show any distinguishable differences on the psychological variables used (MMPI), the demographic variables of Education and Occupational level did differentiate the Stayers from the Non-stayers. When patients who were rated improved in psychotherapy were compared with those rated unimproved, patients with higher occupational achievement and less psychopathology proved to be more successful in treatment.

Received April 1, 1957.

References

1. Auld, F., Jr., & Myers, J. K. Contributions to a theory for selecting psychotherapy patients. *J. consult. Psychol.*, 1954, **18**, 56–60.
2. Barron, F. Some test correlates of response to psychotherapy. *J. consult. Psychol.*, 1953, **17**, 235–241.
3. Barron, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, **17**, 327–333.
4. Eells, K., Davis, A., Havighurst, R. J., Herrick, V. E., & Tyler, R. W. *Intelligence and cultural differences: a study of cultural learning and problem solving.* Chicago: Univer. of Chicago Press, 1951.
5. Gallagher, J. J. MMPI changes concomitant with client-centered therapy. *J. consult. Psychol.*, 1953, **17**, 334–342.
6. Gough, H. G. A new dimension of status: I. Development of a personality scale. *Amer. sociol. Rev.*, 1948, **13**, 401–409.
7. Gough, H. G. A nonintellectual intelligence test. *J. consult. Psychol.*, 1953, **17**, 242–246.

8. Hollingshead, A. B., & Redlich, F. C. Schizophrenia and social structure. *Amer. J. Psychiat.*, 1954, **110**, 695–701.
9. Redlich, F. C., Hollingshead, A. B., Roberts, B. H., Robinson, H. A., Freedman, L. Z., & Myers, J. K. Social structure and psychiatric disorders. *Amer. J. Psychiat.*, 1953, **109**, 729, 734.
10. Roberts, B. H., & Myers, J. K. Religion, national origin, immigration, and mental illness. *Amer. J. Psychiat.*, 1954, **110**, 759–764.
11. Warner, W. L., Meeker, Marchia, & Eells, K. *Social class in America*. Chicago: Science Research Ass., 1949.
12. Welsh, G. S. Factor dimensions A and R. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. of Minnesota Press, 1956. Pp. 264–282.
13. Windle, C. Psychological tests in psychopathological prognosis. *Psychol. Bull.*, 1952, **49**, 451–482.

Journal of Consulting Psychology
Vol. 22, No. 1, 1958

Analysis of Self and Peer Personality Ratings of Psychotherapists and Comparison with Patient Ratings[1]

Eli A. Rubinstein
Veterans Administration and Catholic University

Studies attempting to compare factor patterns of normal and maladjusted persons derived from personality measurements on the same battery of tests have been relatively few. Among these studies the usual finding has been that the factor structure is similar even though the levels of scores differ.

As part of a larger research on evaluation of change during psychotherapy, 100 psychotherapists in ten VA mental hygiene clinics anonymously rated themselves on a multidimensional rating scale previously used with a patient sample. At the same time, 100 peers made ratings of these same therapists, with self raters sometimes acting as peer raters. These data lend themselves to further examination of the relationship between factor patterns of normal individuals versus a patient population.

The major hypothesis tested was that the personality patterns of a nonpatient sample would not differ from personality patterns emerging from analysis of a patient sample. The analysis of the ratings was in terms of 25 items representing six factors previously identified and confirmed with two successive samples of neuropsychiatric patients. Tetrachoric correlations of the 25 items with each other were computed for the 100 self ratings and the 100 peer ratings. The resulting two matrices were each independently clustered according to these previous factor patterns and a multiple group method of factoring was employed. For the peer ratings of therapists, this a priori clustering did reveal linearly independent groups. Rotation to simple structure produced essentially the same six factor patterns as originally isolated with the patient samples. For the self ratings of therapists, factoring of the a priori clusters broke down because of the linear dependence of one of the groups.

As a result of this latter finding, the comparable pairs of correlations in the two matrices of ratings were inspected. Evidence for greater generalization in peer ratings is to be found in the fact that only 35 self rating correlations are greater than .40, while 74 of the peer rating correlations exceed that value. Chi square for the two complete sets of correlations, split at a common median, is 9.63, $p < .01$.

In further examination of the two matrices, 20 pairs of correlations were found in which the significance of the difference between the pair was beyond the .01 level. In all but two cases the high correlation was in the matrix of peer ratings. In all but three of these twenty pairs of correlations, one or both of the scales are rated, for the peer ratings, mainly on inference. Apparently, there is less concomitance of certain manifest and covert behavior, as reflected by their self ratings, than would be presumed from peer ratings.

The finding suggests that when comparisons of ratings on covert behavior are made between self and other ratings, interpretation of similarities or differences should be made with considerable caution.

Brief Report.
Received October 4, 1957.

---

[1] An extended report of this study may be obtained without charge from Eli A. Rubinstein, Neuropsychiatric Research Laboratory, Veterans Administration, Veterans Benefits Office, Munitions Building, Washington 25, D. C., or for a fee from the American Documentation Institute. Order Document No. 5432, remitting $1.25 for microfilm or $1.25 for photocopies.

Journal of Consulting Psychology
Vol. 22, No. 1, 1958

A Scale for Personality Rigidity[1, 2]

John M. Rehfisch
Berkeley, California

This is a report of the construction, by empirical methodology, of a scale for personality rigidity.

As part of an earlier project (6), a survey was made of a number of different types of personality studies in order to define the major elements and correlates of the rigidity–flexibility dimension. The types of behavior referred to by the dimension were found to be quite similar among all the studies, and the following summary of the qualities most commonly included at the rigid pole was made: (*a*) constriction and inhibition, (*b*) conservatism, (*c*) intolerance of disorder and ambiguity, (*d*) obsessional and perseverative tendencies, (*e*) social introversion, (*f*) anxiety and guilt.

This summary may be considered as a comprehensive connotative definition of the term personality rigidity, established empirically according to the criterion of common psychological usage. The synopsis, though more extensive, incorporates all the elements of Fisher's summary description of personality rigidity which is the following: "flexibility, fluidity, rigidity, perseveration and obsessive-compulsiveness . . . have all tended to refer to the idea that the behavior of people in different situations varies along a continuum marked at one extreme by cautious guardedness or limitation of reaction, and at the other extreme by freedom of reaction or lack of defensive caution" (2, p. 342).

The degree to which the scale assesses the characteristics denoted by the above two synopses will be considered as an indication of its validity as a measure of personality rigidity.

Procedure and Results

Setting

The raw data used in this study had been gathered previously in the course of a number of assessment and field-testing programs carried on by the University of California's Institute of Personality Assessment and Research (IPAR) (5).

Subjects

The assessed *S*s whose responses were used in constructing the scale included the following: (*a*) 80 advanced graduate students (40 in each assessed sample) from various departments in the University of California, (*b*) 80 senior medical students (40 in each assessed sample) from the University of California Medical School, (*c*) 70 candidates for admission to the University of California Medical School, and (*d*) 100 Air Force captains. The total N was therefore 330.

---

[1] This paper presents a continuation of a study, the first stage of which is contained in a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree in the department of psychology, University of California, Berkeley. The author is very much indebted to Harrison Gough for his advice and encouragement throughout the course of this study and for his critical reading of the manuscript; and to Donald MacKinnon, Director of the Institute of Personality Assessment and Research, for many valuable suggestions and for permission to use the data and facilities of the Institute.

[2] A portion of this research was supported in part by the USAF under Contract No. AF 18(600)–8, monitored by Technical Director, Detachment No. 7 (Officer Education Research Laboratory), Air Force Personnel and Training Research Center, Maxwell Air Force Base, Alabama. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government. Personal views or opinions expressed or implied in this publication are not to be construed as necessarily carrying the official sanction of the Department of the Air Force or of the Air Research and Development Command.

Bureau Ednl. Psy. Res[illegible]
DAVID HA[illegible] [illegible]AINING COLLEGE
Dated

Criterion Ratings

As part of the assessment program, each assessee was rated by IPAR staff members, consisting of professional psychologists and graduate research assistants, on a number of variables, including that of rigidity. The mean rigidity rating for each *S* served as the criterion for scale derivation.

The number of raters varied from sample to sample, but there were never fewer than five, and the average for all the groups was eight.

The interrater reliability of the average rating, computed by an analysis of variance technique (1), ranged, across the samples, from .50 to .81, with an overall average of .73. Rigidity apparently exists as a sufficiently perceptible personality trait, so as to be rateable with significant interrater agreement.

Item Selection

The scale was compiled from a pool of 957 "true"–"false" personality inventory items taken from the MMPI, the California Psychological Inventory (CPI) (4), and a number of items especially constructed for the IPAR assessments. The construction of the scale involved item analyses in which the inventory results of the highest and lowest 25 per cent on rated rigidity were compared and items chosen according to the significance level of the difference between the proportions of "true" responses among the highs and the lows.

Cross-validation

An unbiased validity measure, for a scale which has been empirically derived from a given sample, can be obtained only from a separate independent sample. Such a requisite for a just validity check is, however, incompatible with the primary objective of obtaining items most stably related to the criterion dimension; for the latter objective requires that all the available criterion samples be used for item selection. This problem was resolved by constructing two preliminary scales from successively larger parts of the total sample, and in both cases employing other portions of the sample for obtaining the required independent validity measures. The obtained measures were then considered as approximate estimates of the validity of the final scale which was derived from the total criterion sample.

The first preliminary scale was based on the responses of the highest and lowest 25 per cent on rated rigidity from among the 80 graduate students, the 80 medical students, and 40 of the 100 AF captains. A cross-validation coefficient for the first scale, computed by correlating scale scores against ratings of rigidity for the 60 AF captains who had not been used in deriving the scale, came to .35 (two-tailed test: $p < .01$). This coefficient of .35 was considered as indicating that the first scale gives an adequate approximation of rigidity, and that this cross check confirmed the validity of the procedures utilized in its derivation.

The second scale was a modification of the first as determined by the responses of the additional 60 AF captains. In the case of this second scale, a more extensive cross-validating appraisal was obtained by correlating scale scores with a large number of traits, as indicated by ratings and Q sorts made by the IPAR staff. The 70 medical school applicants, none of whom had been used in deriving the second scale, were employed for this analysis.

The scale was found to correlate positively with the following traits: (*a*) in interpersonal situations tends to be a listener or spectator .54; (*b*) constriction .51; (*c*) tends toward over-control of his needs and impulses . . . delays gratification unnecessarily .44; (*d*) manifest anxiety .39; (*e*) has a readiness to feel guilty .36; (*f*) tends to delay or avoid decision .35; (*g*) does not vary roles; relates to everyone in the same way .33; (*h*) tends to ruminate and have obsessive thoughts .30; (*i*) is uncomfortable with uncertainties and complexities .30; (*j*) follows routine in living; is orderly .28; (*k*) rigidity .19.

Negative correlations are with: (*a*) is self-indulgent − .51, (*b*) fluency of ideas − .45, (*c*) verbal fluency − .40, (*d*) impulsivity − .35, (*e*) originality − .35.

A comparison of the above findings with the previously presented synopses of personality rigidity reveals an appropriate congruence between the correlates of the second scale and most of the elements contained in or derivable from the synopses. The high scoring *S*s tend to be rated as constricted, over-controlled, anxious, socially introverted, guilt prone, inflexible in their social roles, compulsively doubtful, obsessive and ruminative, uncomfortable with uncertainties and complexities, and orderly. Two of the highest correlates, constriction .51 and over-control .44, are, in psychological parlance, practically synonymous with the term "personality rigidity."

By contrast, low scoring *S*s are likely to be judged as outgoing in social situations, fluent in thought and speech, impulsive, original, and self-indulgent.

These characteristics of low scorers are clearly indicative of flexibility, the bipolar opposite of rigidity.

The low correlation with the rating of rigidity could be due primarily either to error in the scale or to error in the ratings. Since, in the present instance, the scale shows substantial and appropriate correlations with the key elements of the rigidity-flexibility dimension, it seems reasonable to question the accuracy and/or comprehensiveness of these particular rigidity ratings.

The likelihood of considerable error in a rating situation does not of course disqualify ratings as useful criteria in personality research; but does indicate the desirability of using a relatively large sample of ratings, as was done in the present study, in order to balance out random error variance.

Further cross-validating data on the second scale was obtained from an item analysis of the staff-checks on the Gough Adjective Check-List (3). As part of the assessment program, staff judges completed check-lists for each assessee. Composite staff descriptions were then derived by considering each adjective which was selected for an *S* by three or more observers as being present, and all others as being absent. For the present analysis, adjective composites for the 18 highest and the 18 lowest scorers on the second rigidity scale from among the 70 medical school applicants were compared. The following adjectives were found to differentiate significantly (.05 level) between high and low scorers.

Adjectives more often checked about high scorers: anxious, conscientious, conservative, deliberate, dependent, gentle, inhibited, mild, moderate, modest, painstaking, peaceable, quiet, reserved, retiring, serious, shy, sincere, submissive, thorough, timid, withdrawn, worrying.

Adjectives more often checked about low scorers: active, adaptable, aggressive, argumentative, assertive, clear-thinking, confident, curious, demanding, energetic, independent, irritable, organized, outgoing outspoken, planful, poised, quick, resourceful, self-centered, self-confident, self-seeking, sociable, spontaneous, talkative, versatile.

The most salient dimension defined by the two sets of adjectives seems to be a bipolar factor characterized by an anxious restraint, and a meticulous conscientiousness, and conservativism at one extreme; as contrasted with a vigorous, versatile, self-confident spontaneity of thought and action at the other. These two contrasting categories are, in turn, very similar to the "cautious guardedness or limitation of reaction" vs. the "freedom of reaction or lack of defensive caution" which, as mentioned earlier, Fisher (2, p. 342) found to be the most essential characteristics of the rigidity-flexibility dimension.

The final scale (to be known as *Ri*) was based on the inventory results of the highest and lowest 25 per cent on rated rigidity—83 highs and 83 lows—from the total criterion sample.[3] *Ri* is composed of 39 items, 20 significant at the .01 level, 18 significant at the .05 level, and 1 significant at the .06 level.

Scale Items

The 39 *Ri* items, and the direction of the rigid response, are listed below. The items are grouped according to psychological categories, subjectively derived from an impressionistic content analysis made by the writer.

Anxiety and constriction in social situations. 1. I usually don't like to talk much unless I am with people I know very well. (T) 2. I like to talk before groups of people. (F) 3. It is hard for me to start a conversation with strangers. (T) 4. I would like to be an actor on the stage or in the movies. (F) 5. It is hard for me to act natural when I am with new people. (T) 6. I feel nervous if I have to meet a lot of people. (T) 7. I usually feel nervous and ill at ease at a formal dance or party. (T) 8. When I work on a committee I like to take charge of things. (F)

---

[3] The number of *S*s per item actually varied somewhat from item to item, since not all the samples took identical inventories. For the 39 *Ri* items, the *N* ("highs" plus "lows") per item ranged from 106 to 166, with a mean of 132.

9. I usually take an active part in the entertainment at parties. (F) 10. I am a better talker than listener. (F) 11. I try to remember good stories to pass them on to other people. (F) 12. I am embarrassed with people I do not know well. (T) 13. A strong person doesn't show his emotions and feelings. (T)

Need for a stable, orderly, predictable environment; perseverative tendencies. 14. I must admit that it makes me angry when other people interfere with my daily activity. (T) 15. I find that a well-ordered mode of life with regular hours is congenial to my temperament. (T) 16. It bothers me when something unexpected interrupts my daily routine. (T) 17. I don't like to undertake any project unless I have a pretty good idea as to how it will turn out. (T) 18. I find it hard to set aside a task that I have undertaken, even for a short time. (T) 19. I don't like things to be uncertain and unpredictable. (T)

Slowness in coming to a decision—compulsive doubting. 20. I am very slow in making up my mind. (T) 21. At times I feel that I can make up my mind with unusually great ease. (F)

Conservatism and conventionality. 22. I must admit I try to see what others think before I take a stand. (T) 23. I do not like to see women smoke. (T) 24. I would be uncomfortable in anything other than fairly conventional dress. (T) 25. I keep out of trouble at all costs. (T) 26. It wouldn't make me nervous if any members of my family got into trouble with the law. (F) 27. I must admit that I would find it hard to have for a close friend a person whose manners or appearance made him somewhat repulsive, no matter how brilliant or kind he might be. (T) 28. I would certainly enjoy beating a crook at his own game. (F) 29. I would like the job of a foreign correspondent for a newspaper. (F)

Self-doubt and sensitivity to negative criticism. 30. I get very tense and anxious when I think other people are disapproving of me. (T) 31. I am certainly lacking in self-confidence. (T) 32. Criticism or scolding makes me very uncomfortable. (T)

Misanthropy and parsimony. 33. Most people inwardly dislike putting themselves out to help other people. (T) 34. I am against giving money to beggars. (T) 35. Many of the girls I knew in college went with a fellow only for what they could get out of him. (T)

Emphatic concern with work and study. 36. I always follow the rule: business before pleasure. (T) 37. I get disgusted with myself when I can't understand some problem in my field, or when I can't seem to make any progress on a research problem. (T)

Miscellaneous. 38. I have never been made especially nervous over trouble that any members of my family have gotten into. (T) 39. I have no fear of spiders. (T)

The items suggest that the *S* rated high on rigidity differs, as indicated by self-report, from the low-rated *S* in being socially and emotionally constricted; anxious; intolerant of disorder, irregularity, and unpredictability; perseverative; slow in making decisions; conservative; conventional; lacking in self-confidence; misanthropic; and obsessionally involved in work.

Table 1

Summary Statistics for the 39 Item *Ri* Scale and for 27 Items of *Ri* Contained in the CPI

Sample	*N*	39 Item Scale M	39 Item Scale *SD*	27 Item Scale M	27 Item Scale *SD*
AF captains	343	16.01	5.01	10.65	4.32
Medical school applicants	70	14.60	5.85	9.13	4.73
Total	413	15.77	5.21	10.39	4.48

Since most of these qualities are indicative of rigidity, it may be concluded that, among the assessed *S*s, self-descriptions tend to be consistent with the rigidity ratings.

Reliability of the Scale

The reliability of *Ri* was estimated by the split-half technique. In order to approximate equivalence of halves, an attempt was made to match items for the two halves according to similarity of content. The corrected reliability came to .72 in a sample of 60 field-tested AF captains. These *S*s were not included among the 100 assessed AF captains used in deriving the scale.

Summary Statistics

Table 1 contains the mean and *SD* for *Ri* and for the 27 items of *Ri* which are contained in the CPI,[4] among a sample of 343 AF captains and a sample of 70 medical school applicants. The mean and *SD* for the combined distributions are the most representative norms currently available.

In conclusion, it should be noted that the present applicability of *Ri* is limited to men by the exclusively male composition of the samples used in scale construction.

---

[4] The 27 CPI items, as numbered in the CPI booklet, and the direction of the rigid response are as follows: 13(T), 14(T), 16(T), 26(T), 38(T), 50(T), 52(F), 58(T), 83(T), 85(T), 109(T), 150(T), 159(T), 177(T), 219(T), 223(T), 239(F), 267(F), 282(T), 284(T), 288(T), 296(F), 305(T), 328(T), 418(T), 456(F), 471(F).

Summary

A summary of the typical connotations of the personality rigidity concept was presented.

By means of an item analysis technique, applied to 957 personality inventory items, a 39-item scale for personality rigidity was derived. The criterion used for scale construction was ratings of rigidity by staff assessors from the Institute of Personality Assessment and Research made on a total of 330 *S*s. The highest and lowest 25 per cent on rated rigidity were used for the item analysis.

Positive cross-validating evidence for two preliminary versions of the scale was presented.

Received April 11, 1957.

References

1. Ebel, R. L. Estimation of the reliability of ratings. *Psychometrika,* 1951, **16**, 407–424.
2. Fisher, S. An overview of trends in research dealing with personality rigidity. *J. Pers.,* 1949, **17**, 342–351.
3. Gough, H. G. Reference handbook for the Gough adjective check-list. Unpublished manuscript, Univer. of California, Berkeley, Institute of Personality Assessment and Research, 1955.
4. Gough, H. G. *The California Psychological Inventory.* Palo Alto, California: Consulting Psychologists Press, 1956.
5. MacKinnon, D. W. The Institute of Personality Assessment and Research. Unpublished manuscript, Univer. of California, Berkeley, Institute of Personality Assessment and Research, 1951.
6. Rehfisch, J. M. Rigidity, character, and temperament: A psychological analysis. Unpublished doctoral dissertation, Univer. of California, 1954.

Journal of Consulting Psychology
Vol. 22, No. 1, 1958

Validity of the Marsh-Hilliard-Liechti MMPI Sexual Deviation Scale in a State Prison Population[1]

John B. Wattron[2]
Texas Prison System

A scale derived from MMPI items for detecting sexual deviation tendencies has been reported by Marsh, Hilliard, and Liechti (1) which selected 88 per cent of sex deviants studied at the expense of 11 per cent false positives among UCLA students. Subsequently, Peek and Storms (2) used this scale in a state hospital, comparing scores of sex offenders, mental patients, and psychiatric aids. They found the original cutting point of 30 described by Marsh *et al.*, to be useless in discriminating between these groups. A new cutting score of 42 was more effective, yet neither point differentiated between sex offenders and mental patients. They concluded that the scale measured gross maladjustment, rather than sexual deviation tendencies, a possibility that the original authors admit.

In the present work, 60 sexual offender inmates were matched with an equal number of other type felons on age, race, and sex. All were administered the full scale MMPI. Mean scores and overlap comparisons were made between these two groups and those of the original study by use of present and previously determined cutting points on the Sexual Deviation scale. Scores of sex offenders on the MMPI scales were compared with their Sexual Deviation scale scores by correlation procedures.

Mean scores for the felon group and the sex offender group were 46.9 (SD, 9.3) and 49.2 (SD, 9.4) respectively, both of which exceed the mean score of 41.8 (SD, 9.1) reported for sex deviants by Marsh *et al.* Previously determined cutting points were ineffective in discriminating between the two groups of prisoners. The best cutting score, namely 48, was inadequate in differentiating between groups according to the chi-square test. Correlations were obtained between Sexual Deviation scale scores and MMPI scales on inmate sex offenders as follows: *Pt*, .63, and *D*, .58 ($p = .001$); *Pd*, .52, *Sc*, .51, and *Ma*, .49 ($p = .01$); and *F*, .35 ($p = .05$).

These results suggest that the Sexual Deviation scale is indeed a measure of gross maladjustment and that this instrument is of no practical value in discriminating between sex offenders and other type felons in correctional settings.

Brief Report.
Received September 20, 1957.

References

1. Marsh, J. T., Hilliard, Jessamine, & Liechti, R. A. A sexual deviation scale for the MMPI. *J. consult. Psychol.*, 1955, **19**, 55–59.
2. Peek, R. M., & Storms, L. H. Validity of the Marsh, Hilliard, and Liechti sexual deviation scale in a state hospital population. *J. consult. Psychol.*, 1956, **20**, 133–136.

---

[1] An extended report of this study may be obtained without charge from John B. Wattron, 806 West 31 St., Austin, Texas, or for a fee from the American Documentation Institute. Order Document No. 5433, remitting $1.25 for microfilm, or $1.25 for photocopies.

[2] Now in residence, University of Texas.

Journal of Consulting Psychology
Vol. 22, No. 1, 1958

The Use of Multivariate Statistical Analysis of Minnesota Multiphasic Personality Inventory Scores in the Classification of Delinquent and Nondelinquent High School Boys

Peter P. Rempel[1]
University Counseling Service, State University of Iowa

The research worker in psychology and education is frequently confronted with the problem of analyzing differences between two or more groups of individuals with respect to several variables. The problem of differentiating certain selected groups in terms of MMPI scores, for instance, is a good case in point. However, the kinds of statistical procedures commonly used may be subject to criticism on several grounds. Briefly, the several empirical methods of approach which have been developed lack the rigidity of a strictly mathematical model, while such univariate procedures as the *t* test or analysis of variance, when applied to separate keys of the MMPI, fail to take into account the amount of correlation that exists between scores on these keys. These procedures, in short, have failed to provide methods for obtaining mathematically rigid statements of profile or pattern differences.

It is one of the purposes of this paper to provide an illustration of the application of more recently developed statistical procedures, namely the generalized distance function (D^2) and Fisher's discriminant function (LDF) to the solution of problems involving the comparison of MMPI profiles. These techniques have been employed to find answers to the following questions:

1. To what extent can groups of nondelinquent and delinquent-prone ninth-grade high school boys be *differentiated* in terms of MMPI profiles alone or in combination with school record information?

2. To what extent can ninth-grade boys be *classified* correctly as either nondelinquents or potential delinquents in terms of MMPI profiles alone or in combination with school record information?

Basic Population

The basic population of this study consists of ninth-grade boys enrolled in the public high school system of the city of Minneapolis for the school year 1947–1948, and to whom the MMPI was administered by Hathaway and Monachesi (4) in the spring of 1948. Of the 2,521 boys enrolled, 1,997 completed the test. Invalid test profiles with *L* scores of 10 or over and/or *F* scores of 15 or over and omissions for various other reasons reduced the number of valid boys' profiles to 1,802.

In January, 1950, the median age of these youngsters tested was 17 years. A check made of the records of the Hennepin County Probation Office and the Juvenile Division of the Minneapolis Police Department revealed that of those 1,802 boys 376 had appeared before the court, the police, or both. A second follow-up in 1952 raised the number of delinquent boys with valid profiles to 714. These profiles were distributed, in terms of severity of misconduct and number of delinquent acts committed, among eight levels of delinquency as shown in Table 1. A preliminary study failed to reveal any statistically significant differences between a randomly selected set of non-

---

[1] Submitted to the graduate school of the University of Minnesota in partial fulfillment of the requirements for the Ph.D. degree.

Table 1
Distribution of Valid MMPI Profiles by Delinquents, Level of Delinquency, and Nondelinquents

Classification	Number	Per Cent (Based on $N=1802$)	Per Cent (Based on $N=714$)
All Delinquents	714	39.62	
Levels 1 and 2	363	20.15	50.84
Level 3	231	12.82	32.35
Level 4	47	2.61	6.58
Level 5	31	1.72	4.34
Level 6	22	1.22	3.08
Level 7	12	0.67	1.68
Level 8	8	0.44	1.12
Total Level 3–8	351	19.47	49.16
Nondelinquents	1,088	60.37	—
Total (all delinquents and nondelinquents)	1,802	100.00	—

delinquent profiles and those of the delinquents assigned to the lower two levels. This left a total of 351 valid delinquent profiles distributed among Levels 3–8.

In the major section of this study, referred to as the main study, this set of 351 delinquent profiles was divided by random sampling procedures into two groups: Group A ($N = 175$) which was used in the development of the classification formula, and Group B ($N = 176$) set aside for cross-validation purposes. Two samples of 175 profiles each were similarly selected from the nondelinquent set of valid profiles. No statistically significant differences, in terms of level of delinquency and educational status achieved by 1952, were found to exist between delinquent samples A and B. The nondelinquent samples were found to be equal in terms of the latter criterion. The number of nondelinquents from a particular school included in samples A or B was made to match the number of delinquents from that school.

An analysis of three variables; namely, family status, mobility rating, and ratings of rent level, by means of chi-square techniques, failed to discriminate between delinquent and nondelinquent samples at statistically signifi-

Table 2
Comparison of Sample A Means and Variances for MMPI Scores of Nondelinquent and Delinquent Groups

MMPI Scales	Total Delinquents ($N = 175$)		Stratified Sample of Nondelinquents ($N = 175$)		Tests of Significance[a]	
	Mean	Variance	Mean	Variance	t	F
L	3.90	4.81	3.81	5.77	.12	1.20
K	13.76	15.91	13.49	17.26	.21	1.08
F	6.90	13.30	5.40	10.53	4.06**	1.26
Hs	52.86	67.98	50.59	86.30	2.42*	1.27
D	52.28	96.19	51.64	96.17	.61	1.00
Hy	52.12	55.51	51.33	71.77	.93	1.29
Pd	64.18	88.45	57.51	110.94	6.23***	1.25
Mf	50.08	62.37	53.58	83.07	3.83***	1.33
Pa	53.52	78.40	50.68	72.72	3.05**	1.08
Pt	57.47	97.11	54.99	85.07	2.43*	1.14
Sc	62.02	119.56	57.05	88.68	4.54**	1.35
Ma	61.17	114.14	56.95	101.14	3.97***	1.13
Sie	56.66	66.83	53.85	63.18	1.38	1.16

Note.—Similar tests of significance using the ratio d^2/S^2 yielded significant differences for the same eight scales.

[a] Level of significance of difference between means and variance estimates has been noted by asterisks:
* Significance at the .05 level;
** Significance at the .01 level;
*** Significance at the .001 level.

Table 3

Comparison of Sample A′ Means and Variances for the Variables of the Delinquent and Nondelinquent Groups in the School Study

Variables	Nondelinquents ($N = 91$)		Delinquents ($N = 91$)		Tests of Significance[a]	
	Mean	Variance	Mean	Variance	t	F
L	3.7912	4.38	3.8022	6.23	.03	1.42
K	13.6813	21.06	13.7373	19.66	.08	1.07
F	5.8462	14.72	7.1758	10.91	2.49*	1.35
Hs	51.5055	60.29	53.3956	72.96	1.55	1.21
D	51.6263	90.94	52.4514	96.50	.06	1.06
Hy	51.6483	49.56	52.0771	54.07	.40	1.09
Pd	57.7912	94.89	64.9011	114.10	4.67**	1.20
Mf	51.9340	81.11	49.0440	80.71	2.16*	1.01
Pa	52.2418	88.38	54.3297	80.43	1.53	1.10
Pt	55.3516	96.92	57.6593	67.01	1.71	1.45*
Sc	57.5824	141.17	62.7912	73.18	3.38**	1.93**
Ma	56.0330	141.72	61.4835	137.01	3.26**	1.19
Si	52.1538	71.30	53.0549	52.71	.50	1.35
Days absent	7.0549	120.75	12.8791	32.76	4.96**	3.68**
IQ	103.4900	104.71	99.6800	130.41	2.36*	1.25
H.P.R.	2.1687	.3228	1.7171	.5574	4.57**	1.73**

[a] The level of significance is indicated by asterisks:
* .05 level of significance;
** .01 level of significance.

cant levels. To the extent that these three factors tend to reflect the socioeconomic status of the student population under discussion, delinquents and nondelinquents may be regarded as coming from essentially the same background.

A comparison of the means and variances for the MMPI scores of delinquents and nondelinquents is presented by means of Table 2. Choice of MMPI variables for inclusion in the classification formula was made on the basis of their potential contribution to the generalized distance function, D^2. One statistic which provides a measure of this contribution is the ratio of the square of the mean group difference to an estimate of the variance of the distribution for each variable, referred to as d^2/S^2, another the measure of correlation between variables. The combination of variables resulting in the widest possible separation of the two groups in terms of D^2 were the scores on the *F*, *Hs*, *Sc*, *Mf*, and *Pd* scales, arranged in that order.

The School Study

A second part of the study, referred to as the school study, dealt with the problem of identifying juvenile delinquents on the basis of a combination of MMPI scores and school record data. The same basis population was used in both studies. However, the group from which experimental samples were chosen in the school study consisted of a total of 869 boys used by Roessel (12) in his study of high school dropouts. Of these, 412 had left high school without graduating, while the rest ($N = 457$) constituted a random sample of those who had graduated. Of these 869 boys, 182 proved to be delinquents with valid MMPI profiles. From the remaining 687 profiles, a control group of 182 nondelinquent profiles was chosen. As in the main study, both delinquent and nondelinquent groups were divided into Samples A′ and B′. Chi-square tests revealed no statistically significant differences in terms of delinquency level and educational status between the delinquent samples of the school and the main studies, nor between the nondelinquent samples of the two studies in terms of the latter criterion. Comparisons of mean scores and variance estimates for Sample A′ are reported by means of Table 3. Variables selected in terms of maximum contributions to D^2 were scores on the *Ma, Sc, Mf,* and *Pd* scales of the MMPI, honor point ratio, and number of days absent, in that order.

Discussion of Techniques Employed

Rao (9) has shown that the degree of overlap between any two clusters of points in a multidimensional space is measured by the generalized distance function, D^2. The greater the numerical value of D^2, the smaller the amount of overlap, so that this measure may be regarded as the "distance" between groups.

To eliminate intervariable correlations Rao (11) suggested a method of transforming the original set of variables to an orthogonal set. Collier (3) presented illustrations of the uses of Rao's pivotal condensation techniques to effect the transformation in the process of computing D^2. Methods for testing the significance of D^2 have also been developed by Rao (8, 9, 10).

The application of Fisher's discriminant function to the problem of classification into one or two known multivariate normal populations was discussed by Anderson (1). Tests of the effectiveness of the LDF in classifying individuals correctly were presented by Lubin (7). For complete and comprehensive discussions of the techniques employed in this study the reader is referred to Christensen (2), Lockman (6), and Johnson (5).

Results

In the main study, analysis of the data yielded a theoretical probability of *correct* classification for both delinquents and nondelinquents of 68.39 per cent. The *overall* percentage is essentially the same as that achieved by chance when the ratio of nondelinquents to delinquents in the population is known to be 80.5 to 19.5 per cent, as it is in this study. However, classification of delinquents by the LDF is more than three times as effective as chance procedures, the formula identifying 68.4 per cent of the delinquents as against 19.5 per cent identified by chance.

Empirical Cross Validation of Samples A and B

The actual effectiveness of the linear discriminant function was tested empirically by applying the formulas developed on Sample A to the delinquent and nondelinquent sub-

Table 4

A Summary Comparison of Classification Efficiency for the LDF Procedure with Chance Expectation for Samples A and B in the Main Study by Delinquent and Nondelinquent Groups within the Sample

Samples Used	Correct Classification by Formula		Correct Classification by Chance		Percentage Improvement over Chance[a]
	N	%	*N*	%	
Sample A (Dels.) (*N* = 175)	120	68.6	87.5	50.0	18.6**
Sample A (Nondel.) (*N* = 175)	115	65.7	87.5	50.0	15.7**
Total Sample A (*N* = 350)	235	67.1	175.0	50.0	17.1**
Sample B (Dels.) (*N* = 176)	124	70.5	88.0	50.0	20.5**
Sample B (Nondel.) (*N* = 175)	103	58.9	87.5	50.0	8.9*
Total Sample B (*N* = 351)	227	64.7	175.5	50.0	14.7**
All Groups (*N* = 701)	462	65.9	350.5	50.0	15.9**

[a] Level of significance is indicated by means of asterisks:
* .05 level of significance;
** .01 level of significance.

Table 5

Classification of Delinquents by Level of Delinquency Using LDF Procedures for Two Randomly Chosen Samples of the Main Study: Sample A on Which Prediction Formulae Were Developed and Sample B on Which the Formulae Were Applied Through Cross Validation

		Percentage of Correct Classification					
		Sample A ($N = 175$)		Sample B ($N = 176$)		Total ($N = 351$)	
Level[a]	Total	N	%	N	%	N	%
3	231	82	70.7	78	67.8	160	69.3
4	47	14	60.9	18	75.0	32	68.1
5	31	7	46.7	12	75.0	19	61.3
6	22	9	81.8	8	73.0	17	77.3
7	12	4	66.7	5	83.3	9	75.0
8	8	4	100.0	3	75.0	7	87.5
Total	351	120	68.6	124	70.5	244	69.5

[a] In addition to the 351 delinquents in Levels 3 to 8 above which had been randomly assigned to Samples A and B, 363 pupils had been classified as delinquents of Levels 1 and 2. Application of the same procedures identified 75 or 42.3% of the Level 1 group as delinquent and 111 or 60.0% of the Level 2 group, an overall correct classification of 51.2%, approximately the same as chance expectation.

samples in both Samples A and B. A study of cross-validation results as presented in Table 4 indicates that the overall percentage of correct classifications, namely 65.9 per cent, fell somewhat short of the predicted percentage of 68.4. The fact that the formula correctly identified only 58.9 per cent of the nondelinquents in Sample B accounts for most of the shrinkage. However, 68.6 per cent of delinquents in Sample A, and 70.5 per cent in Sample B were correctly identified.

Classification results were also broken down in terms of delinquency level, as presented in Table 5. The tabulation shows generally increasing proportions of correct classification with level of delinquency, Level 5 being the only exception.

Results of the School Study

The probability of *correct* classification for both delinquents and nondelinquents was found to be 70.9 per cent.

Table 6

A Summary Comparison of Classification Efficiency for the LDF Procedures with Chance Expectations for Samples A′ and B′ in the School Study by Delinquent and Nondelinquent Groups within the Sample

	Correct Classification by Formula		Correct Classification by Chance		Percentage Improvement over Chance[a]
Samples Used	N	%	N	%	
Sample A′ (Dels.) ($N = 91$)	60	65.9	45.5	50.0	15.9**
Sample A′ (Nondels.) ($N = 91$)	72	79.1	45.5	50.0	29.1**
Total Sample A′ ($N = 182$)	132	72.5	91.0	50.0	22.5**
Sample B′ (Dels.) ($N = 91$)	63	69.2	45.5	50.0	19.2**
Sample B′ (Nondels.) ($N = 91$)	63	69.2	45.5	50.0	19.2**
Total Sample B′ ($N = 182$)	126	69.2	91.0	50.0	19.2**
Total Samples A′ and B′ ($N = 364$)	258	70.9	182.0	50.0	20.9**

[a] The level of significance is indicated by means of asterisks. Thus, two asterisks indicate significance at the .01 level.

9.70
291

Table 7

Classification of Delinquents by Level of Delinquency Using LDF Procedures for Two Samples of the School Study: Sample A′ on Which Prediction Formulas Were Developed and Sample B′ on Which the Formulas Were Applied Through Cross Validation

Level	N	Sample A′ (N = 91)		Sample B′ (N = 91)		Total (N = 182)	
		N	%	N	%	N	%
3	106	38	71.70	31	59.62	69	65.71
4	30	9	60.00	12	75.00	21	67.74
5	21	9	75.00	6	66.67	15	71.43
6	13	5	71.43	5	83.33	10	76.97
7	7	1	50.00	3	60.00	4	57.14
8	5	1	50.00	3	100.00	4	80.00
Total	182	63	69.23	60	65.93	123	67.58

Results of cross-validation procedures (see Table 6) indicate that classification by the LDF permitted the correct identification of 65.9 per cent of the delinquents in Sample A′ and of 69.2 per cent in Sample B′. Corresponding figures for nondelinquents are 79.1 per cent and 69.2 per cent for Samples A′ and B′ respectively. In each case improvement over chance is significant at the .01 level.

Although the two studies are not strictly comparable, the addition of school record data raised the *overall* predictive efficiency of the LDF from 65.9 per cent in the main study to 70.9 per cent in the school study, a difference significant at the .05 level.

Classification results in terms of the level of delinquency are presented by means of Table 7. Again, an increase in the proportion of correct classifications corresponds closely with increase in level of delinquency, a marked drop in predictive efficiency at Level 7 being the only exception in the general trend.

Discussion

Unlike a number of empirical methods which have been suggested for the comparison of multiphasic profiles, the evidence presented here suggests that multivariate statistical techniques provide a rigid measure of the distance between two groups in terms of *all* the available numerical information. Unlike the measures achieved as a result of the application of univariate procedures to separate scales individually, the generalized distance function provides a single unit of measurement which is communicable, and lends itself particularly to the purpose of making comparisons from one study to another. Unlike univariate procedures, multivariate techniques may be used for the purposes of classification and prediction.

The relatively high percentage of correct classifications achieved in this study, in spite of the well-known high degree of intercorrelation of MMPI variables [2] and the numerically small average group differences on individual variables, may serve as an indication of the even greater usefulness of these techniques in research studies where these limiting factors do not exist to the extent they were found in the present investigation.

Summary

This study has demonstrated the usefulness of applying multivariate statistical techniques to the analysis of MMPI scale scores alone, or in combination with school data, for the purpose of classifying ninth-grade boys as potential delinquents or nondelinquents. The techniques employed proved to be effective to the extent that 62.3 per cent of the nondelinquents and 69.5 per cent of the delinquents were correctly identified by the use of multiphasic data alone; while a combination of

---

[2] Thus the median product-moment correlation coefficients for the five variables of Sample A in the main study was .30, with a range of .06 to .49; corresponding figures for the six variables of Sample A′ of the school study were: median .04; range − .28 to .36.

multiphasic and school record data made possible the correct identification of 74.2 per cent of nondelinquent boys and 67.5 per cent of delinquent-prone boys.

Received March 22, 1957.

References

1. Anderson, T. W. Classification by multivariate analysis. *Psychometrika,* 1951, **16**, 31–50.
2. Christensen, C. M. Multivariate statistical analysis of differences between pre-professional groups of college students. *J. exp. Educ.,* 1953, **21**, 221–232.
3. Collier, R. O. Some application of the method of pivotal condensation in statistical analysis. *J. exp. Educ.,* 1953, **21**, 233–241.
4. Hathaway, S. R., & Monachesi, E. D. *Analyzing and predicting juvenile delinquency with the MMPI.* Minneapolis: Univer. of Minnesota Press, 1953.
5. Johnson, M. C. Classification by multivariate analysis with objectives of minimizing risk, minimizing maximum risk, and minimizing probability of misclassification. *J. exp. Educ.,* 1955, **23**, 259–264.
6. Lockman, R. A. An evaluation of naval aviation cadet selection measures using multivariate discriminatory statistical techniques. Unpublished doctoral dissertation, Univer. of Minnesota, 1954.
7. Lubin, A. Linear and nonlinear discriminating functions. *British J. Psychol., statist. Sec.,* 1950, **3**, 90–103.
8. Mahalanobis, P. C., Majundar, D. N., & Rao, R. C. Anthropometric survey of the United Provinces, 1941: a statistical study. *Sankhya,* 1949, **9**, 89–324.
9. Rao, C. R. The problem of classification and distance between two populations. *Nature,* 1947, **160**, 835–836.
10. Rao, C. R. Tests of significance in multivariate analysis. *Biometrika,* 1948, **35**, 58–79.
11. Rao, C. R. *Advanced statistical methods in biometric research.* New York: Wiley, 1952.
12. Roessel, F. P. Minnesota Multiphasic Personality Inventory results for high school drop-outs and graduates. Unpublished doctoral dissertation, Univer. of Minnesota, 1954.

Journal of Consulting Psychology
Vol. 22, No. 1, 1958

Progressive Matrices (1938) and Emotional Disturbance[1]

Sidney Kasper
Evansville State Hospital

The Progressive Matrices (PM), unlike other indices of intelligence, has to date received little attention with essentially psychotic populations. An instrument intended as a measure of "clarity of thinking" and "reasoning ability" might legitimately be expected, in the presence of illness, to reveal certain functional differences. Sensitivities of this type are thought evincible in the PM when contrasted with a verbal test relatively resistant to the effects of disturbance, such as vocabulary. This paper summarizes: (*a*) PM and the Wechsler Adult Intelligence Scale Vocabulary test (WAIS *V*) relationships to morbidity scores derived from the Lorr Multidimensional Rating Scale for Psychiatric Patients (MSRPP) and (*b*) the relationship to morbidity of two points or more deviation between expected and actual scores in any of PM's five sets of 12 items each. These deviations have been thought to possess "psychological significance."

Subjects included 50 consecutive first admissions to Evansville State Hospital who had not yet attained their 56th year, who showed no apparent organic involvements, and were capable of complying with instructions.[2] WAIS *V* scores were prorated and then grouped so as to fall within one of three categories: (*a*) below 90 IQ, (*b*) 90–110 IQ, and (*c*) above 110 IQ. PM percentile scores for the same subjects were converted and grouped similarly.

PM and WAIS *V* are in 48% agreement. Their differences do not permit rejection of the null hypothesis and may be attributable to chance. Contrary to prediction, 82% ($N = 41$) of the cases may be described as scoring as high or higher on the PM than on WAIS *V*, with PM scores being greater in 34% ($N = 17$) of the total cases. These findings are not attributable to differences in age or educational attainment and are at variance with the expected trend. The .50 correlation between PM and WAIS *V* ($p < .05$) is essentially similar to those of previous investigations using similar measures with less severely disturbed samples or with normals.

No meaningful relationship is found between ratings of morbidity and estimates of intellectual functioning. Differences in mean morbidity scores for the various discrepancies between intelligence measures do not clearly demonstrate a concomitant variation between events. Similarly, with respect to the specific MSRPP scales, "conceptual disorganization" and "perceptual distortion," nonsignificant changes in these morbidity scores are associated with differences between intellectual measures. The results, then, consistently fail to demonstrate PM's efficiency in discriminating the influence of pathology.

Finally, none of the disparities among PM sets of scores is meaningfully related to morbidity score. Therefore, clinical use of this feature of the test would seem severely limited.

The results, in general, point to considerable ambiguity and lack of validity in interpretations of PM scores with heterogeneous clinical populations.

Brief Report.
Received October 14, 1957.

---

[1] An extended report of this study may be obtained without charge from Sidney Kasper, Evansville State Hospital, Evansville, Indiana, or for a fee from the American Documentation Institute. Order Document No. 5431, remitting $1.25 for microfilm or $1.25 for photocopies.

[2] The author is especially indebted to Don R. Brown for the statistical analysis of the data.

Journal of Consulting Psychology
Vol. 22, No. 1, 1958

The Polarity of Psychological Tests

Marshall B. Jones[1]
US Naval School of Aviation Medicine, Pensacola, Florida

A psychological test may be named after either of its extremes. The same test, for example, might be named either "Rigidity" or "Flexibility," depending upon which end of the continuum is to be emphasized. Which of the two names we use is commonly considered a matter of convention. Implicit in this convention is the assumption that both ends of the continuum are equally meaningful. Conceptually, however, the dimension we intend to measure may be very far from meeting this assumption. "Anxiety" is a good example. We recognize immediately that the idea of anxiety is far clearer than is the idea of nonanxiety. The anxious end of the conceptual continuum is more meaningful or, if you like, has a more consistent meaning than has the nonanxious end. If this difference between the two ends of the conceptual continuum is reflected in the test itself, the homogeneity of items falling at one end of the continuum will be greater than the homogeneity of items falling at the other end. Put more specifically, the average intercorrelation among difficult[2] items will be greater (or less) than the average intercorrelation among easy items. To the extent that there is a relationship between item difficulty and average interitem correlation we will speak of the test as *polarized*. The purpose of this paper is briefly to consider the scope and consequences of test polarity.

An Illustration

A discussion of polarity is perhaps best introduced by an illustration. The Pensacola Z Survey (4) contains five scales. Four of these correspond to familiar clinical ideas: Anxiety, Dependency, Rigidity, and Hostility.[3] Each of the four scales contains forty items in forced-choice form, with no item overlap between scales. All four scales are substantially homogeneous (Hoyt coefficients ranging from .82 to .90) and consist of items which most psychologists would recognize as related to the trait the scale is intended to measure. All four scales are scored in the neurotic direction. Thus, a high score indicates Anxiety, Dependency, Rigidity, or Hostility. The distributions of the four scales divide them into two pairs: Anxiety and Hostility, which are skewed positively, and Dependency and Rigidity, which are skewed negatively. To illustrate polarity the items of each scale were broken down into four groups according to difficulty level. Because of the differences in skew between the Anxiety-Hostility pair and the Dependency-Rigidity pair the same categories would not do for both pairs. To 16 of the Anxiety items, for example, and to 12 of the Hostility items, less than 30% of the subjects responded in the neurotic direction, while to none of the Dependency and to only three of the Rigidity items did less than 30% respond in the neurotic direction. For Anxiety and Hostility, therefore, the items were divided into items to which less than 30% of the subjects responded in the neurotic direction, between 30% and 40%, between 40% and 50%, and over 50%. For Dependency

---

[1] Opinions or conclusions contained in this paper are those of the author. They are not to be construed as necessarily reflecting the view or the endorsement of the Navy Department.

[2] Throughout this paper we shall use the terminology of aptitude testing in referring to the proportion of persons responding with the key to a given item. The terminology of "difficult" and "easy" items sounds strange in the context of personality testing, but there does not seem to be any equally well-understood alternative.

[3] The Z survey contains a fifth scale, called Heteronomy, which is not included here as its items overlap with all four of the other scales.

Table 1

Average Tetrachoric Correlations Between Z-Survey Items, Grouped by Scale and Difficulty Level, Together with the Numbers of Items Involved

Scale	$p \leq .3$	$.3 < p \leq .4$	$.4 < p \leq .5$	$.5 < p$
Anxiety	.389 (16)	.314 (11)	.235 (9)	.195 (4)
Hostility	.221 (12)	.196 (8)	.145 (8)	.145 (12)
	$p \leq .5$	$.5 < p \leq .6$	$.6 < p \leq .7$	$.7 < p$
Dependency	.173 (6)	.173 (5)	.208 (12)	.222 (17)
Rigidity	.244 (11)	.242 (8)	.235 (11)	.313 (10)

and Rigidity, the items were grouped according as the proportion of subjects responding in the neurotic direction was less than 50%, between 50% and 60%, between 60% and 70%, and over 70%. Tetrachoric correlations were available between all items of each scale. The average correlation within each group of items was then obtained. In Table 1 these averages are presented together with the number of items in the group. A glance at Table 1 shows that polarity is present.[4] In Anxiety, for example, the smaller the proportion of people who respond in the anxious direction the higher the homogeneity. To a lesser degree the same trend obtains for Hostility. The neurotic ends of these two scales, Anxiety and Hostility, are most consistent. Precisely the reverse obtains with respect to Dependency and Rigidity. The *larger* the proportion of persons responding in the dependent direction, the higher the homogeneity of the items. For these two scales, independence and flexibility are the more consistent ends of the scale, though for Dependency the relationship is not strong. Polarity, then, definitely does obtain, at least in the scales of the Z survey. However, the nature of the Z survey and its construction lead us to believe that very much the same results would be obtained with other personality questionnaires.

Homogeneity

To the extent that a test is polarized we may expect the homogeneity of the test to be greater (or less) in a high-scoring sample of subjects than in a low-scoring sample. This proposition asserts the principal pragmatic consequence of test polarity. To establish its truth let the total score T_i for subject i be divided into two components X_i and Y_i, where X_i is the subject's score on the easier half of the items and Y_i is his score on the harder half. In addition, consider two samples of subjects: a high-scoring sample and a low-scoring sample. Most of the subjects in the high-scoring sample will answer most of the easy items correctly. Therefore, for the high-scoring group the standard deviation, S_X, will be small in comparison to S_Y. In the low-scoring sample, however, most of the subjects will answer most of the hard items incorrectly. Therefore, for them S_Y will be small in comparison to S_X. Our proposition asserts that if the test is polarized the "homogeneity" of the test in the high-scoring sample will be different from its homogeneity in the low-scoring sample. To come to grips with this proposition it is necessary to be more specific about homogeneity, which means, essentially, what Cronbach (1) has called equivalence, "the degree to which the test score indicates the status of the individual at the present instant in the general and group factors defined by the test." There is a fairly wide choice of coefficients: any one of the Guttman lower bounds (2), including the Kuder-Richardson (7), an analysis of variance coefficient (3), or Loevinger's homogeneity coefficient (8). The proposition at issue could be established with any one of these. However, it is most easily established with the Kuder-Richardson formula which is at the same time the most familiar of the available coefficients.

In the low-scoring sample let $S_X = KS_Y$, where $K > 1$. For the sake of simplicity we will assume that this relationship is exactly reversed in the high-scoring sample, i.e., that $S_Y = KS_X$. In both samples the Kuder-Richardson coefficient is given by

---

[4] As has already been mentioned, Anxiety and Hostility are skewed positively, i.e., toward anxiety and hostility, while Dependency and Rigidity are skewed negatively, i.e., toward independence and flexibility. In all four instances, therefore, the direction of skew lies toward the more polarized end of the scale. This relationship is not accidental. There is an algebraic relationship between polarity and skewing. However, other variables, e.g., average item difficulty, affect skewing as well as polarity. A variable might well be skewed and entirely unpolarized at the same time. However, to the extent that a test is polarized it will tend also to be skewed. Thus, though the two phenomena cannot be equated, the existence of skew constitutes good reason to look for polarity.

$$r_{TT} = \frac{2N}{2N-1}\left[1 - \frac{\sum_g S_g^2}{S_T^2}\right],$$

where $2N$ is the number of items in the test (both halves) and where g ranges over all $2N$ items. By the algebraic manipulation of the formula for r_{TT} we find that

$$r_{TT}=\frac{2N}{2N-1}\left[\frac{K^2\left(\frac{N-1}{N}\right)r_{XX}+\left(\frac{N-1}{N}\right)r_{YY}+2Kr_{XY}}{1+K^2+2Kr_{XY}}\right]$$

in the low-scoring sample and that

$$r_{TT}=\frac{2N}{2N-1}\left[\frac{\left(\frac{N-1}{N}\right)r_{XX}+K^2\left(\frac{N-1}{N}\right)r_{YY}+2Kr_{XY}}{1+K^2+2Kr_{XY}}\right]$$

in the high-scoring sample, where r_{XX} and r_{YY} are the Kuder-Richardson coefficients for X and Y respectively and where r_{XY} is the product-moment coefficient between X and Y.[5] As is clear from these two equations, if r_{XX} and r_{YY} are equal the homogeneity coefficients in the two samples are identical. In this context, equality of r_{XX} and r_{YY} is tantamount to no polarity. Therefore, if there is no polarity, the two homogeneity coefficients are equal. On the other hand, if $r_{XX} \neq r_{YY}$, polarity exists. In terms of the present model degree of polarity is given by the difference, $r_{XX} - r_{YY}$. To the extent that r_{XX} is greater than r_{YY}, the homogeneity of the test in the low-scoring sample will be greater than its homogeneity in the high-scoring sample; and conversely. If a test is polarized toward anxiety, for example, the difference between two anxious subjects will be more reliable than a difference of the same magnitude between two nonanxious subjects. The implications for tests which are used or intended to be used in both clinical and normal populations are clear. If the test is polarized, its internal properties cannot remain the same in both populations. If track were kept of polarity, we might be able to account for some of the instances in which a clinically developed instrument "falls apart" in a normal population and vice versa.

Significance

In order to deal effectively with polarity, one must be able to say whether the polarity of a given test is statistically significant or not. The means about to be described for meeting this requirement are manifestly imperfect. However, the problem at issue is not simple; there may be no practical yet logically rigorous and conceptually adequate way of assessing the significance of polarity.

Table 1 expresses the notion of polarity in terms of the average tetrachoric correlation within each of four groups of items, the items being arranged in order of difficulty. To obtain a coefficient from these data one needs only to calculate a rank correlation coefficient (6, 9) between the average tetrachoric[6] correlation within a group and the difficulty level of the group. However, with only four groups much in the way of significance is not likely to be established. The computational load, moreover, is considerable. In Table 2 the results of breaking the scales of the Z survey into groups of three items each are set forth. There are 13 groups for each scale, beginning with the three most difficult items and ending with the three least difficult. In all

---

[5] In these formulas r_{XX} is assumed to have the same value in the low-scoring sample as in the high-scoring sample; in a similar way, r_{YY} is supposed to remain constant in the two samples. These assumptions are tantamount to supposing that X and Y are not themselves polarized. However, if the entire test $(X + Y)$ is polarized, its parts, X and Y, will very probably be polarized as well. By assuming that X and Y are not polarized we minimize the effects of polarity on homogeneity. Polarization within the two halves as well as between them would simply reinforce the effects on homogeneity of the difference in polarity between the hard and the easy halves of the test.

[6] Throughout this paper tetrachoric coefficients are used. Our use of the tetrachoric in preference to the phi coefficient needs some justification. The sampling distribution of the phi coefficient is not known. However, whatever the distribution may be, its expected value must be dependent upon item difficulty. If we sample from a population with a known value of phi, the expected value of phi in the sample will depend upon the item difficulties of the items involved. The idea of polarity, however, turns entirely on the relationship between item difficulty and inter-item correlation. In consequence, the use of the phi coefficient would completely confound polarity with statistical artifact. On the other hand, the expected value of the tetrachoric coefficient is independent of item difficulty.

Table 2

Average Tetrachoric Correlations Within Groups of Three Items Arranged in Order of Difficulty, Together with the Corresponding Polarity Coefficients, for the Four Z-Survey Scales

Scale	Item Group													
	1	2	3	4	5	6	7	8	9	10	11	12	13	τ
Anxiety	41	40	57	17	47	24	27	35	22	23	21	13	19	.54
Hostility	29	14	12	25	35	21	11	15	25	16	04	15	28	.14
Dependency	16	17	12	17	18	01	28	20	06	08	34	31	36	−.39
Rigidity	38	20	24	17	30	28	27	20	33	24	29	25	52	−.18

four scales the item which ranked 19th was dropped. As is clear from the table, the results yield very much the same picture as did the results set forth in Table 1. However, the number of tetrachorics which had to be calculated for each scale has been reduced from around 200 to 39. At the same time the n available for a test of significance has increased from four to thirteen. However, even with groups of three items each the n for the test of significance is small. In Table 2 the rank correlations (tau) for the four scales appear. Of these, only the coefficient for Anxiety is significant. Fortunately, the n can be considerably increased. It is not necessary to treat the entire, undivided sample. The total sample may be divided into subsamples, and a coefficient calculated which the writer has called "a partial rank correlation coefficient" (5) and which Torgerson has called "a rank correlation coefficient within subgroups" (10). In general, the value of the partial coefficient will approximate the value obtained with the whole sample. However, the standard error of the partial coefficient will equal exactly $(k)^{-\frac{1}{2}}$ times the standard error of the original coefficient, where k is the number of subsamples. The use of subsamples has the effect of simply increasing the n upon which the test of significance is based. The number of subsamples into which the total may be broken is limited primarily by the nature of the tetrachoric coefficient. So many subsamples may not be used that one or more of the cells involved in calculating the tetrachorics became empty, since in that circumstance the tetrachoric would not be defined. In practice, the actual subsampling is not needed. The data of Tables 1 and 2, for example, are based on 310 subjects and to no item does a majority greater than .85 respond in the same way. Therefore, the sample may be separated into five subsamples of 62 subjects each with some confidence that the tetrachorics will retain their meaning. The standard error for the original taus was .21; therefore, the standard error for the partial taus will be .115. Since the values of the partial tau will approximate the values based on the entire sample, Dependency is definitely polarized. Hostility and Rigidity, however, would seem not on the evidence presented to have been established as polarized. For most purposes this general method, though far from elegant, seems to give an adequate evaluation of polarity.

Summary

Many psychological traits are polarized in the sense that one end of the conceptual continuum is clearer, i.e., has a more consistent meaning, than the other. In the first section of this paper, the four peripheral scales of the Pensacola Z Survey were all found to involve some degree of polarity. Polarization of the scale, Anxiety, was particularly powerful. The construction and content of the Z survey is, moreover, quite typical of personality inventories in general. It is, therefore, reasonable to suppose that many, if not most, personality scales will be found to reflect the polarity of the conceptual continuum they are supposed to measure. The primary significance of polarity is that to the extent that a test is polarized the homogeneity of the test in high-scoring samples will be greater (or less) than its homogeneity in low-scoring samples. In the second section of this paper this result was established. In the third and last section, a procedure was proposed by which the degree of polarity a test exhibits can be tested for significance. The procedure involves the calculation of Kendall's tau between the average tetrachoric correlation within groups of items and the difficulty of the item group in which the tetrachoric average obtains. The tau is then tested for significance.

Received March 21, 1957.

References

1. Cronbach, L. J. Test "reliability": Its meaning and determination. *Psychometrika,* 1947, **12,** 1–16.
2. Guttman, L. A basis for analyzing test-retest reliability. *Psychometrika,* 1945, **10,** 255–282.
3. Hoyt, C. Test reliability obtained by analysis of variance. *Psychometrika,* 1941, **6,** 153–160.
4. Jones, M. B. The Pensacola Z Survey: A study in the measurement of authoritarian tendency. *Psychol. Monogr.,* in press.
5. Jones, M. B. An addition to Schaeffer and Levitt's "Kendall's Tau." *Psychol. Bull.,* 1957, **54,** 159–160.
6. Kendall, M. G. *Rank correlation methods.* London: Griffin, 1948.
7. Kuder, G. F., & Richardson, M. W. The theory of the estimation of test reliability. *Psychometrika,* 1937, **2,** 135–138.
8. Loevinger, Jane. A systematic approach to the construction and evaluation of tests of ability. *Psychol. Monogr.,* 1947, **61,** No. 4 (Whole No. 285).
9. Schaeffer, M. S., & Levitt, E. E. Concerning Kendall's Tau, a non-parametric correlation coefficient. *Psychol Bull.,* 1956, **53,** 338–346.
10. Torgerson, W. S. A non-parametric test of correlation using rank orders within subgroups. *Psychometrika,* 1956, **21,** 145–152.

Journal of Consulting Psychology
Vol. 22, No. 1, 1958

An Experimental Group Version for School Children of the Progressive Matrices[1]

Read D. Tuddenham, Louis Davis, Leslie Davison,
and Richard Schindler

University of California

An experimental group test version of Raven's Progressive Matrices was prepared by the "ditto" reproduction process, to explore the suitability of the test for American grade-school children. Features of the experimental version were elimination of the separate answer sheets and of the color printing used in Raven's 1947 edition. Two classrooms, selected to represent different socioeconomic levels, were tested at each grade level from third through sixth.

Mean scores showed the expected progression from grade to grade. Reliability coefficients for separate grade levels range from .87 to .94.

Tentative percentile norms were presented for each grade level. Despite the relative crudity of our dittoed version and the small samples tested, the norms are very close to those presented by Raven for Scottish children. Our data do not show the consistent superiority of American children reported by Green and Ewert (1).

Percentages passing per item are reported separately by grade. While the actual difficulty values differ from those of Green and Ewert, the order of items by difficulty corresponds very closely to their findings. It shows marked consistency from grade to grade within our own data but differs appreciably from the published sequence.

Differences between boys and girls in mean scores and in standard deviations for separate grade levels are inconsistent and very small in magnitude. Sex differences appear not to be related to test performance.

Differences between upper-middle-class and working-class schools are very substantial and highly significant. Despite the novel content of Progressive Matrices and its freedom from dependence on language, the test reveals marked differences between children from different socioeconomic levels.

Correlations between Progressive Matrices and Kuhlmann-Anderson, California Mental Maturity, and Lorge-Thorndike tests are similar in magnitude, clustering around .40 or .45, despite an interval of as long as two years between the administration of the two tests correlated. Green and Ewart report similar values.

In conclusion, the Progressive Matrices seem to be well adapted to group testing in elementary schools. It is brief, reliable, and capable of engaging the interest and attention of school children of the grade levels investigated. Our results indicate that a black and white version is satisfactory for children as young as third graders, and marking answers on the booklets instead of on separate answer sheets avoids a possibly more serious source of confusion than any which might result from the absence of color. In view of the great potential usefulness of a nonlanguage test at once so brief, reliable, and broad in range of difficulty, it is to be hoped that Progressive Matrices will soon be made available in an inexpensive "throw away" booklet form.

Brief Report.
Received October 21, 1957.

Reference

1. Green, M. W., & Ewert, Josephine C. Normative data on Progressive Matrices (1947). *J. consult. Psychol.*, 1955, **19**, 139–142.

---

[1] An extended report of this study may be obtained without charge from R. D. Tuddenham, Dept. of Psychology, University of California, Berkeley 4, Calif., or for a fee from the American Documentation Institute. Order Document No. 5429, remitting $1.75 for microfilm or $2.50 for photocopies.

Journal of Consulting Psychology
Vol. 22, No. 1, 1958

Test Anxiety Level and Goal-Setting Behavior

E. Philip Trapp and Donald H. Kausler
University of Arkansas

In recent years, the effects of anxiety as a drive state have been studied on a diversity of activities, ranging from simple motor tasks to complex verbal problems (e.g., 1, 3, 4, 7, 9). These studies have been concerned primarily with the facilitating or inhibiting effects of anxiety on performance level. A related problem, which has not been specifically investigated, is the effect of test anxiety on goal-setting behavior.

This study is primarily concerned with an aspect of the relationship between anxiety and aspiration level. More exactly, the study involves a comparison of level of aspiration scores (LOA scores) between *S*s with high test anxiety and low test anxiety over several trials on a learning task. The problem may be phrased in terms of the null hypothesis, i.e., there is no difference between the LOA scores of a high test anxiety group and a low test anxiety group.

Method

Eighty-four students, approximately the same number of each sex in two sections of an undergraduate course in abnormal psychology at the University of Arkansas, were given the test anxiety questionnaire developed by Sarason and Mandler (2, 5, 6). The questionnaire was given during regular class periods. The *S*s were told that the purpose was to measure the attitude of students at the university toward taking various types of tests. The instructions, administration, and scoring of the questionnaire followed the procedure outlined by Sarason and Mandler. The 28 *S*s scoring in the upper third for this distribution of scores were arbitrarily designated as the high test anxiety group (Group HA); the 28 *S*s scoring in the lower third were arbitrarily designated as the low test anxiety group (Group LA).

One week later, the *S*s were given an extended group form of the Wechsler-Bellevue Digit Symbol Test (8). The general instructions read as follows:

> The task you are about to do is the digit symbol subtest of the Wechsler-Bellevue Intelligence Scale. This subtest predicts very successfully total IQ when it is administered in an individual setting. We are interested in seeing if it can predict as well when administered in a group setting. Therefore, it is of particular importance that you listen carefully to the instructions and follow them accurately.
>
> It has been found that people either overestimate considerably or underestimate considerably their performance on an ego-involved task such as an intelligence test. Consequently, you will be asked to predict how well you think you will do before taking the subtest.
>
> This study is being repeated in other universities throughout the country to see if there are sectional differences both in respect to performance on the digit symbol test and on ability to predict performance. Therefore, make an all-out effort to perform at your best.

The *S*s were then presented with a test booklet. The front page asked for identification data, such as name, age, etc. Also, space was provided for *S* to record both his estimate of his future performance (level of aspiration) and his actual performance score (level of performance) on each of the four trials he received in the test period. The other four pages of the booklet were identical. Each page contained a copy of the Wechsler-Bellevue Digit Symbol Test, increased in length to 75 unfilled boxes. The standardized instructions for the subtest were read to the *S*s, and the sample set of problems completed. The *S*s were then asked to turn back to the front page of the booklet and record the number of boxes they thought they could fill in a minute and a half. After each trial, the *S*s counted their number of completions and recorded the total beneath their prediction score. They then

Table 1

Means and Standard Deviations of Groups HA and LA for LOA and LOP Scores by Trials

Group	Score	Trial 1 M	Trial 1 *SD*	Trial 2 M	Trial 2 *SD*	Trial 3 M	Trial 3 *SD*	Trial 4 M	Trial 4 *SD*
HA	LOA	55.21	8.54	57.00	9.73	62.14	8.63	65.07	8.72
	LOP	53.79	9.35	59.61	9.44	64.32	9.40	—	—
LA	LOA	56.43	7.34	60.43	6.64	66.50	5.95	69.57	6.64
	LOP	56.14	9.45	63.25	7.74	66.64	10.16	—	—

made their estimate for the next trial. Time between trials was two minutes. Apparently, none of the *S*s was aware of any connection between the test anxiety questionnaire and this testing session.

Results

The means and standard deviations of the level of aspiration scores (LOA) and the level of performance scores (LOP) for Group HA and Group LA on successive trials are given in Table 1. Since many *S*s in both groups reached maximum performance on the fourth trial, the performance scores for this trial were highly skewed and, consequently, have been omitted from further analysis.

The difference between mean LOP scores for the two groups on each trial was tested for significance by the *t* test. The obtained values of *t* are given in Table 2. Inspection of Table 2 shows that no significant differences in mean LOP between the two groups occurred on any of the trials. As an additional measure of performance level, a composite mean LOP score was determined for each group, and the difference between group means was tested by the *t* test. This score was found by averaging the three performance scores for each *S* and then determining the mean of these averages. For Group HA, the mean was 59.25 and standard deviation was 9.01. For Group LA, the mean was 62.01 and standard deviation was 8.34. The obtained *t* of 1.37 is not significant at the .05 level of confidence.

Either by treating the trials separately or in combination, the results support the null hypothesis of no real difference in performance ability between the two groups on the digit symbol test. These findings, incidentally, agree, in general, with the results obtained by Mandler and Sarason (2) on comparable groups also given the Wechsler-Bellevue Digit Symbol Test.

The difference between mean LOA scores for the two groups was tested for significance on each trial by the *t* test. The resulting values of *t* are given in Table 2. On the first trial there was no statistically significant difference in LOA between the two groups. This was the only trial in which the *S*s expressed their aspirations without prior direct knowledge of their performance level. Hence, at the start of the testing session there was no real difference in the goal-setting behavior of the two groups.

With knowledge of results commencing on the second trial, an immediate trend can be detected which reaches statistical significance in the third and fourth trials. The personal frame of reference for evaluating aspiration level based on performance information apparently had a differential effect on the goal-

Table 2

Values of *t* for Differences Between Group Means on LOA and LOP Scores

Score	*t* Trial 1	*t* Trial 2	*t* Trial 3	*t* Trial 4
LOA	.56	1.51	2.16*	2.13*
LOP	.92	1.55	.87	—

* Significant at the .05 level.

setting behavior of the two groups. Group LA consistently displayed a higher level of aspiration than Group HA.

Further evidence demonstrating the tendency of Group LA to set a higher level of aspiration following knowledge of results is revealed in a comparison of composite group means. To obtain these means, the aspiration scores of each *S* for the three trials (second, third, and fourth) were averaged and the mean of these averages determined for both groups. For Group HA, the mean was 61.40 and the standard deviation was 8.29. For Group LA, the mean was 65.50 and the standard deviation was 5.63. The obtained *t* of 2.12 between the means is significant beyond the .05 level of confidence.

Discussion

The most striking finding in our study was the increasing divergence in goal-setting behavior between the two groups on successive trials despite correspondingly undifferentiated levels of performance and a roughly equivalent pretest level of aspiration. The differential LOA between the two groups reached statistical significance on the third and fourth trials. This marked tendency of Group LA to express higher goal-setting behavior could lead to several plausible interpretations. The interpretation favored by the authors may be stated as follows. The high test anxious *S*, while objectively performing no poorer, is more condemning and critical of his test behavior than the low test anxious *S*. In the evaluation of his test performance, he then tends to overemphasize the negative factors, is less optimistic over future performance, and expresses this judgment in a relatively lowered LOA. As he gains more information or knowledge of results, his frame of reference becomes increasingly slanted in the conservative direction. The low test anxious *S*, on the other hand, is subjectively more influenced by his success or progress, and is thereby less restrained in his goal-setting behavior.

It should be noted that both groups functioned within the realistic range throughout the testing. The self-devaluating tendency of Group HA did not reach excessive proportions because the task was apparently not overly frustrating. A more difficult task or even the same task given under conditions of high stress would be predicted to bring on a pronounced drop in adaptive behavior and a concomitant expression of an unrealistic level of aspiration in *S*s with high test anxiety. The finding that our two groups did not differ significantly in level of performance is consistent with other reported studies in which anxiety was the independent variable in learning tasks. In a number of related studies, the Iowa group has shown that high anxious *S*s perform relatively better on simple tasks and low anxious *S*s better on complex tasks. The digit symbol test, being of medium difficulty, should mask any differences between the two groups. Both Mandler and Sarason (2) and our study confirmed this expectation.

Summary

Eighty-four college students were administered the test anxiety questionnaire developed by Mandler and Sarason. They were split into two extreme groups on the basis of their scores on the questionnaire. The first group (Group HA) consisted of *S*s scoring in the top third of the distribution. The second group (Group LA) consisted of *S*s scoring in the bottom third. Both groups were given four trials on an extended group form of the Wechsler digit symbol subtest. Prior to each trial the *S*s expressed their level of aspiration (LOA), and after each trial they were given knowledge of their own results. A comparison of mean LOA scores between the two groups was made on each trial, with *t*'s of .56, 1.51, 2.16, and 2.13, respectively. The last two trials showed significant differences, with Group LA reflecting higher LOA scores. A comparison of LOA scores between the two groups was made by combining the data on the last three trials, with the obtained *t* of 2.12 being significant beyond the .05 level of confidence. The mean performance scores between the two groups were compared for each trial, with *t*'s of .92, 1.55, .87, respectively. None of the trials showed significant differences. When the performance data on all of the trials were combined for each group and then compared, a *t* value of 1.37 was obtained, which also was not statistically significant.

Received March 15, 1957.

References

1. Farber, I. E., & Spence, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.,* 1953, **45**, 120–125.
2. Mandler, G., & Sarason, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.,* 1952, **47**, 166–173.
3. Matarazzo, J. D., Ulett, G. A., & Saslow, G. Human maze performance as a function of increasing levels of anxiety. *J. gen. Psychol.,* 1955, **53**, 79–96.
4. Montague, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.,* 1953, **45**, 91–98.
5. Sarason, S. B., & Gordon, E. M. The test anxiety questionnaire: scoring norms. *J. abnorm. soc. Psychol.,* 1953, **48**, 447–448.
6. Sarason, S. B., & Mandler, G. Some correlates of test anxiety. *J. abnorm. soc. Psychol.,* 1952, **47**, 810–817.
7. Spence, K. W., Farber, I. E., & McFann, H. H. The relation of anxiety (drive) level to performance in competitional and noncompetitional paired associates learning. *J. exp. Psychol.,* 1956, **52**, 296–305.
8. Wechsler, D. *The measurement of adult intelligence.* (3rd ed.) Baltimore: Williams & Wilkins, 1944.
9. Wesley, E. Perseverative behavior in a concept-formation task as a function of manifest anxiety and rigidity. *J. abnorm. soc. Psychol.,* 1953, **48**, 129–134.

Journal of Consulting Psychology
Vol. 22, No. 1, 1958

Further Normative Data on the Progressive Matrices[1]

Gerald Sperrazzo[2] and Walter L. Wilkins
Saint Louis University

The normative data on the colored Progressive Matrices (1) suggest that clinicians should be cautious in the application of Raven's original norms, based on Dumfries, Scotland children (4), to American children. The data from the Rochester, Minnesota, children reported by Green and Ewert (1) were in some respects quite different from the Scottish data and led to the inference that there may have been a notable difference in the intelligence level of children in the Scottish and Minnesota schools, or that there was a difference in the socioeconomic levels of the pupils of the normative sample, or that unidentified factors were producing the difference. Prompted by the implication that there were differences in productivity on the Progressive Matrices in the two samples, we tried to cross validate the two sets of normative data with the testing of St. Louis school children. As an additional feature of the study, the selection of the sample was further arranged to allow racial background of the subjects to affect the results, if it should be a relevant factor.

Procedure and Subjects

Except for the gross size of the sample, replication of the Rochester sample for age and grade level was attempted. From the public schools of St. Louis a sample of 480 children was selected and arranged by grade and age as shown in Table 1. Sampling took account of the fact that St. Louis, sometimes described as a border city, contains schools with various compositions of Negro and Caucasian pupils. Approximately one-third of the children came from a school with an all white pupil population, one-third from a school with an all Negro population, and one-third from a school with 60% Negro and 40% white population. As a partial control of economic factors, the occupations of the pupils' fathers were classified within three levels from the Dictionary of Occupational Titles here called: A, professional or semi-professional; B, skilled, semi-skilled, or clerical; and C, service, unskilled, or labor. For each of these three levels, 180 pupils were chosen. The nature of the sample, then, permitted the identification of the source of variance from four variables: age, sex, race, and socioeconomic status.

The Progressive Matrices test was administered in both individual and group form. The book form of the test was used for individual administration with children between the ages of seven and eight. For the older children, 2 × 2 color slides were made of each matrix of the test and projected upon a screen to groups ranging in size from 15 to 30. The group method of administration has been reported previously and found to be reliable with no significant differences between group and individual scores (1).

Results and Discussion

While the data upon Minnesota children suggested that the Scottish norms were low, the results of the 1956 St. Louis testing suggest that they may be closer to the American standard than suspected. Scores reported by Green and Ewert for Rochester children were consistently higher than Raven's Dumfries scores by two to six points throughout the range reported. In St. Louis, however, this apparent American superiority is greatly lessened. The scores reported in Table 1 show St. Louis medians approximating Raven's me-

---

[1] Supported in part by a grant from the Saint Louis University Human Relations Center.

[2] Now at the University of Kentucky.

Table 1

Colored Progressive Matrices Scores for St. Louis Children as Compared with Dumfries and Rochester Children

Age	*N*	Range	Mean	*SD*	Medians		
					St. Louis	Dumfries	Rochester
7/0–7/5	81	4–33	17.27	3.74	16	16	21
7/6–7/11	41	9–30	18.31	5.01	18	17	21
8/0–8/5	69	9–33	20.06	5.53	20	18	24
8/6–8/11	45	7–34	22.44	5.02	23	20	25
9/0–9/5	51	9–34	23.08	5.76	23	22	26
9/6–9/11	50	8–36	23.38	7.87	24	24	27
01/0–10/5	44	10–35	25.20	6.42	26	24	28
10/6–10/11	29	10–34	24.10	9.44	26	26	29
11/0–11/5	70	10–36	27.00	5.48	28	28	30

dians, with the St. Louis children slightly superior at the younger ages and possibly slightly inferior at the upper age ranges.

Inspection of Table 1 indicates that the median scores found through this study agree more consistently with Raven's median scores than with those from the Rochester study. Although the range of difference between the median scores of Raven and those of this study is from zero to three points, the average median score difference is one point. It appears, then, that the results found here support the validity of Raven's norms on the Dumfries school children for application to the St. Louis population.

Accounting for all the variables which may have led to such a notable discrepancy between the median scores of the Rochester study and the present investigation is rather difficult. The socioeconomic status of the sample used in Rochester may be relatively higher than it would be in Dumfries. If this were the case, then an increase in the median scores of the Minnesota children would be expected since the present study yielded evidence that the socioeconomic status of the subjects can significantly affect the results.

Table 1 also indicates that the ceiling of the test was reached by two age groups. Since this has occurred previously in other studies (1), it seems rather clear that the ceiling of the Progressive Matrices test is apparently too low for a common normal pupil population, and a clear differentiation of scores at the upper end of the distribution is lacking.

An analysis of variance was performed to study the operation of the age, sex, race, and socioeconomic status variables. As a matter of convenience the *S*s were placed into five age groups ranging from seven to eleven years of age. Prior to performing the analysis of variance, Bartlett's test of homogeneity was applied to the data and no significance was found. The results of the analysis are reported in Table 2.

Table 2

Analysis of Variance of Scores from 480 Children on Colored Progressive Matrices Test (1947)

Source of Variation	Degrees of Freedom	Mean Square	*F*
Total	479		
Age	4	1193.15	46.62*
Sex	1	80.03	3.13
Race	1	2116.80	82.72*
Socioeconomic (S-E)	2	528.26	20.64*
Age × Sex	4	30.62	1.97
Age × Race	4	33.60	1.31
Age × S-E	8	24.65	
Sex × Race	1	.21	
Sex × S-E	2	62.76	2.45
Race × S-E	2	210.99	8.25*
Age × Sex × Race	4	4.32	
Age × Sex × S-E	8	13.73	
Age × Race × S-E	8	44.26	1.73
Sex × Race × S-E	2	45.01	1.76
Race × Sex × Age × S-E	8	51.65	2.02**
Within groups	420	25.59	

* .01 level of confidence.
** .05 level of confidence.

Differences significant beyond the .01 level of confidence were found with the age, race, and socioeconomic status variables. There was no significant difference between the sexes. The interaction between race and socioeconomic status was also significant beyond the .01 level of confidence, while the third order interaction was found to be significant only at the .05 level.

The significance found in this analysis merits some discussion. The test was built to yield scores increasing with the *S*'s age; therefore, the significant differences found here with age levels require no further justification. On the other hand, the test was not designed to discriminate racial differences. It is apparent from the significant race by socioeconomic status interaction and the third-order interaction that a restriction on the interpretation of the race difference found is necessary. The measured differences in scores between races are related to the age, sex, and socioeconomic status of the *S*s. The results cannot be interpreted, therefore, as showing differences in intelligence between the races tested here. The differences found seem to depend upon variations of the nonrace factors.

Product-moment correlation coefficients were computed between the Progressive Matrices and three standard intelligence tests, utilizing the scores from recent school testing of the pupils. With the Otis Quick-Scoring Intelligence Test the r was .23 (N of 36), with the California Test of Mental Maturity it was .30 (N of 137), and with the Kuhlmann-Anderson it was .41 (N of 215).

Summary and Conclusions

The Colored Progressive Matrices Test was administered to 480 St. Louis school children, ranging in age from seven to eleven years. The subjects were selected from three St. Louis public schools. One was an all-Negro school, another an all-white school, and the third enrolled both white and Negro children. Three levels of socioeconomic status, based upon the occupation of the father, were established for purposes of comparison. The first category included professional and semi-professional workers; the second, skilled, semi-skilled and clerical workers; and the third, laborers, unskilled, and service workers.

The test was administered individually in book form to seven- and eight-year-old children. In the group administration, 2 × 2 color slides were projected to 9- to 11-year-old subjects in groups of 15 to 30.

The findings lead to the following conclusions.

1. The norms established by J. C. Raven for the Progressive Matrices appear to be valid for application to the population represented by the sample in the present study.

2. There is some indication on the basis of the range of scores, that the ceiling of the test is too low for a population of average children.

3. Differences significant beyond the .01 levels of confidence were found between the various age levels (seven to eleven), between races (white and Negro), and between socioeconomic levels. There was no significant difference between the sexes.

4. The Progressive Matrices Test showed a correlation with the Otis Quick-Scoring test of .23, with the California Test of Mental Maturity, .30, with the Kuhlmann-Anderson, .40.

Received March 25, 1957.

References

1. Green, M. W., & Ewert, Josephine C. Normative data on the Progressive Matrices (1947). *J. consult. Psychol.*, 1955, 19, 139–142.
2. Kuhlmann, F., & Anderson, Rose G. *Kuhlmann-Anderson Intelligence Test.* Minneapolis: Educational Test Bureau, 1952.
3. Otis, A. S. *Quick Scoring Mental Ability Tests.* Yonkers, N. Y.: World Book Co., 1952.
4. Raven, J. C. *Progressive Matrices (1947), Sets A, Ab, B.* London: H. K. Lewis, 1947.
5. Sullivan, E. T., Clark, W. W., & Tiegs, E. W. *California Test of Mental Maturity.* Los Angeles: California Test Bureau, 1951.

Journal of Consulting Psychology
Vol. 22, No. 1, 1958

The Achievement Motive and Field Independence[1]

Jack Wertheim and Sarnoff A. Mednick
Harvard University

McClelland *et al.* ascribe the origins of the achievement motive to the "different experiences which children have" (2, p. 288). More specifically, they state that "mothers of sons with low need achievement tend to demand less in the way of independent achievement at an early age, and tend to be more restrictive than other mothers" (2, p. 302). An interesting parallel may be seen in Witkin's description of the developmental precursors of field dependence; he states "field dependent children have been subjected during growth to strongly restrictive parental pressures" (4, p. 87). He also points out that these parents tend not to demand independent behavior from their children.

The elements of similarity in these writers' descriptions of the developmental origins of these two personality variables suggested that an individual's scores on the tests of these factors should be related. More particularly, it was predicted that there is a positive correlation between field independence and need achievement.

Method

Forty-two undergraduate *S*s (31 female, 11 male) were tested for need achievement (n-Ach) and field independence (FI) in that order. *S*s took the n-Ach test in groups of three each. They were shown four slides projected on a screen for 20 sec. each and were asked to answer questions about the slide. As a representative of the FI test battery, a short form (1) of the Embedded Figures Test (EFT) was administered. *S* was shown a complex geometric figure followed by a simple figure and then asked to find the simple figure in the complex figure.

Detailed descriptions of procedure and materials relative to these tests may be found elsewhere (1, 2, 3).

Results and Discussion

The scoring of the n-Ach material was done by a trained scorer whose scoring reliability has been in the .90's in previous tests. The higher an *S*'s score, the greater his n-Ach. The scores ranged from -4 to $+13$, with a mean of 3.4. *S*'s score on the EFT was his total time to locate the embedded figures on all 12 cards. In this case, the lower *S*'s score, the greater his field independence. The scores ranged from 120 to 955 sec., with a mean of 462.8 sec.

The product-moment correlation between n-Ach and EFT scores was $-.40$ ($p < .01$), indicating a significant relationship between the two variables; the higher the achievement motive, the greater the field independence.

It is interesting to note that the similarity in the descriptions of the developmental origins of these two personality traits is reflected in the observed relationship of the variables in adults.

Received June 17, 1957.

References

1. Jackson, D. N. A short form of Witkin's embedded figures test. *J. abnorm. soc. Psychol.*, 1956, 53, 254–255.
2. McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive.* New York: Appleton-Century, 1953.
3. Witkin, H. A. Individual differences in ease of perception of embedded figures. *J. Pers.*, 1950, 19, 1–15.
4. Witkin, H. A. *et al.* A study of change in perception during psychological development and of the relation between changes in mode of perception and changes in personality functioning. *Progress Rep. USPHS Grant No. M-628*, Coll. of Med., State University of New York, 1956.
5. Witkin, H. A. *et al. Personality through perception.* New York: Harper, 1954.

---

[1] The help of Drs. McClelland and Witkin in supplying materials and assistance is appreciated.

Journal of Consulting Psychology
Vol. 22, No. 1, 1958

Assimilation, Failure-Avoidance, and Anxiety[1]

David E. Hunt
Yale University

and Harold M. Schroder
Princeton University

The aims of the present investigation were (*a*) to develop a scale measuring the *tendency to assimilate negative information* and (*b*) to study the relationship of this assimilation tendency to other personality measures and to behavior in a problem-solving situation.

In an earlier series of investigations (4) it was found that a subject's (*S*'s) behavior in an actual situation of failure was related to his interpretation of failure as measured by a paper-and-pencil "situational interpretation" test. In this test, *S* was confronted by a series of hypothetical negative situations (failure or criticism) and instructed to select one of two alternative interpretations following each situation: one alternative represented self-negation or decreased self-evaluation (e.g., "This shows I might not be so good") and the other alternative represented an opposing interpretation of self-affirmation without any change in self-evaluation (e.g., "This is too difficult for anyone to do"). The latter alternative was descriptively regarded as a "failure-avoidant" interpretation because of its implied refusal to accept the possibility of failure. The behavioral measure involved *S*'s attempts to solve a problem under increasing failure stimulation. The tendency for *S* to "leave the field" by selecting a different goal when failing to reach the initial goal was regarded as "failure-avoidant" behavior.

*S*s utilizing failure-avoidant interpretations on the situational test were found to be much more likely to utilize failure-avoidant behavior in the problem-solving situation. In discussing this relationship, it was noted that the failure-avoidant interpretation process may effectively prevent the occurrence of assimilation; that is, in order to be assimilated, a negative event must first be accepted, at least provisionally. An individual experiencing a failing score or a critical remark must first categorize the event as possibly true before he can undertake any adaptive modification. He must say, in effect, "I may be wrong" before he can ask, "How can I correct myself?" The provisional acceptance of negative information and the subsequent self-modification are both viewed as aspects of the assimilation process.

In previous work we have assumed that avoidance and assimilation represent opposing poles on a process continuum but have not verified this assumption by establishing its relationship to other measures. Therefore, in the present investigation a direct measure of the assimilation tendency was developed in order to study the assimilation characteristics, both general and specific, of groups distinguished in terms of certain personality and behavioral variables. Following leads from the findings of other investigators (2), the relationship between anxiety and assimilation was also explored. The following hypotheses concerning general tendencies to assimilate negative informations were tested:

---

[1] The work described in this paper was supported in part by a research grant M-955 from the National Institute of Mental Health of the National Institutes of Health, Public Health Service, and in part under Contract Nonr-1858 (12) sponsored by the group psychology branch, Office of Naval Research. We are greatly indebted to the staff members of the participating schools and camps in Connecticut, New York, New Jersey, and Pennsylvania for their generous cooperation. We would like to express our appreciation especially to Ronald S. Wilson for his help in coding responses and analyzing data, and also to John McDavid, Jr. for his assistance in collecting data.

I. Persons who give avoidant interpretations score lower in assimilation tendency than persons who give nonavoidant interpretations.

II. Persons high in anxiety score lower in assimilation tendency than persons low in anxiety.

III. Persons who show failure-avoidant behavior score lower in assimilation tendency than persons who show nonavoidant behavior.

Method

The assimilation scores were derived from sentence completion responses given to negative stems. Combining the advantages of structured materials with freedom of response, the Sentence Completion Method (SCM) appeared most appropriate for present purposes. According to the provisions for scoring, two individuals may receive the same assimilation score by giving quite different responses, especially in nonassimilation where the range of defensive possibilities such as aggression, denial, etc., is quite varied. In addition to providing an over-all score for general assimilation tendency, the SCM responses could also be later reconsidered to note any specific features of assimilation or nonassimilation which might characterize the groups under consideration.

Subjects and Procedure

The SCM was administered to approximately 800 white boys between the ages of 13 and 18 during 1955. Groups of *S*s ranging in size from five to 80 took the SCM (without any time limit), followed by the Situational Interpretation Experiment (4) and a short verbal intelligence test (6) consisting of 20 vocabulary items. These 800 *S*s will be considered according to the following four samples:

Sample A: 500 *S*s drawn from YMCA camps, Boy Scout camps, public and private schools.

Sample B: 79 *S*s attending a parochial summer school for purposes of raising their grades.

Sample C: 60 *S*s representing the entire junior and senior class in a consolidated high school.

Sample D: 160 *S*s drawn from sources similar to Sample A.

For certain cross-validation purposes, SCM responses of selected *S*s from 880 boys tested during 1956 were used (Sample E).

Materials

The five sentence stems, representing hypothetical situations of criticism and failure, are listed below. The number indicates the position of the stem in relation to the total 30 stems which each *S* completed.

Criticism:

4. When someone criticizes me
19. If someone says I'm wrong
29. If someone says I have done poorly

Failure:

9. When I fail
22. When things go wrong

Development of Assimilation Scale

Obtaining an assimilation score from five SCM responses required three steps: coding each response according to an objective system; transforming each coded response into an assimilation score; and summing the five scores to yield a total assimilation measure.

Coding system. The system developed for coding SCM responses objectively was intended to minimize the number of categories and maximize the amount of information retained. The categories, given below in the Assimilation scale, were described in a manual with examples for each category. Agreement between three judges [2] in coding 260 responses was 93 per cent. In coded form, the responses could not only be easily transformed into an assimilation scale, but were "recoverable" for analysis at a more specific level than assimilation.

Assimilation scale. Each coded response was assigned an assimilation score according to the following three-point scale, in terms of the explicit statement or implied likelihood that *S* would show assimilation or integrated modification in his behavior as a result of the hypothetical negative event described in the sentence stem.

---

[2] Ronald S. Wilson and Roger C. Porr served as judges along with one of the authors, and their assistance is gratefully acknowledged.

Scale Score 0: Any form of denial; negative feeling or aggression; interaction rejected; pessimistic perception of interaction; reciprocation; self-disclaimed or other attributed causality; cessation of, disruption of, or decreased effort in behavior.

Scale Score 1: Interaction accepted or questioned; optimistic perception of interaction; attention to interaction; conforming reaction to other; cognitive activity; behavior repeated; effort maintained.

Scale Score 2: Self-attributed causality; behavioral modification through either compensation or correction; effort increased.

Transformation from coded responses to assimilation scale could be accomplished with almost perfect judge agreement. When the separate scores for the five SCM items were summed, an assimilation scale score (ranging from 0 to 10) was obtained for each *S*. The mean assimilation scale score for 107 randomly selected *S*s in Sample A was 4.76, SD, 2.34. Assimilation scores were not related to scores of verbal intelligence.

Reliability. Two SCM forms were administered to Sample C at a six-month interval. Since the forms changed slightly, there were only three identical negative stems occurring on both forms ("When someone criticizes me," "When I fail," and "When things go wrong"). Computation of six-point assimilation scores based on these three stems for both forms yielded a test-retest reliability coefficient of .64 ($N = 48$), which was estimated (by Spearman-Brown prophecy formula) to be .75 if based on five, rather than three, stems.

Identification of Extreme Groups

Failure-avoidant interpretation. This variable was defined by the score on the criticism-failure subscale of the Situational Interpretation Experiment (4). Each of the ten items on this subscale presented a hypothetical situation of criticism or failure with instructions for *S* to select one of the two alternative choices that would best indicate the thought he would have in that situation. The total number of self-negating or "nonavoidant" choices was added to obtain *S*'s score so that a low score indicated a greater tendency to make avoidant interpretations. From Sample A, a failure-avoidant interpretation group (scoring 0 to 2) of 19 *S*s and a nonavoidant interpretation group (scoring 10) of 31 *S*s were selected (see Table 3 in reference 4 for distribution of scores). Elimination of one *S* from each group because of an incomplete SCM left 18 and 30 *S*s, respectively, or approximately the lower four per cent and the upper six per cent of the distribution.

Manifest anxiety. This variable was defined by the score on the Taylor Manifest Anxiety Scale (5). From Sample B, a low anxiety group (scoring 2 to 8) of 11 *S*s and a high anxiety group (scoring 28 and above) of 11 *S*s were selected representing the lower and upper 13 per cent of the distribution.

Failure-avoidant behavior. This variable was defined as the extent to which *S* avoids a situation in which he is receiving gradually increasing failure by selecting a different goal rather than continuing to strive toward the same goal (4). From Sample C, a behavioral failure-avoidant group of 14 *S*s who avoided the situation at the earliest signs of failure and a behavioral nonavoidant group of 14 *S*s who remained in the situation through all 12 trials were selected, representing the upper and lower 29 per cent of the distribution.

Results

Assimilation Scale

Hypothesis I. Failure-avoidant interpretation. As indicated in Table 1, the mean assimilation score for the avoidant interpretation group is significantly lower ($< .01$) than

Table 1

Assimilation Score Comparisons of Extreme Groups

Hypothesis	Extreme Group	Assimilation Score Mean	*t*	*df*	*p**
I	Failure-avoidant interpretation	3.06	2.99	46	<.01
	Nonavoidant interpretation	5.17			
II	High anxiety	2.27	3.80	20	<.01
	Low anxiety	5.45			
III	Failure-avoidant behavior	4.01	1.61	26	<.10
	Nonavoidant behavior	5.94			

* By one-tailed test.

the mean for the nonavoidant group. Extreme groups on this interpretation measure were selected from Sample D in an attempt at replication. In this sample, the mean of 21 avoidant *S*s scoring 0–3 (lower 9 per cent) was lower, but not significantly so ($p < .10$, one-tailed test), than the mean of the 27 nonavoidant *S*s scoring 9–10 (upper 18 per cent) ($t = 1.55$, $df = 46$). A nonsignificant Pearsonian r of .12 was obtained between these two measures on 107 randomly selected *S*s, indicating that whatever relationship exists is limited to extreme groups.

Hypothesis II. Anxiety. The mean assimilation score for the high in manifest anxiety group is significantly lower ($< .01$) than the mean for the group low in anxiety. This inverse relationship is also supported by significant correlations in the samples: $r = -.32$ for 79 *S*s in Sample B ($< .01$) and $r = -.28$ for 32 *S*s in Sample A for whom anxiety scores were available ($< .05$).

Hypothesis III. Failure-avoidant behavior. The hypothesis concerning avoidant behavior was not supported, although the difference between assimilation scores was in the predicted direction ($p < .10$). A test of this hypothesis in another sample also failed to confirm it.

Specific Assimilation Characteristics of Extreme Groups

The coded SCM responses were examined for more specific characteristics of assimilation or nonassimilation in each group, and two SCM code categories, denial and aggression, were identified as potentially differentiating. The denial code included the negation of feeling ("I don't mind") and negation of aggression ("I don't get mad") as well as denial of attention to the source of negative information ("I ignore them"). The aggression code included explicitly hostile responses (e.g., "I get mad," "I blow up"). These measures were recorded simply in terms of occurrence or nonoccurrence on any one of the five SCM stems. Incidence of denial and aggression was 17 and 14 per cent, respectively, for 107 randomly selected *S*s.

In addition to these two SCM measures, an index of the incidence of self-correction was also available for the extreme groups differentiated with respect to failure-avoidant behavior. A situation of academic failure was described and *S* was instructed to imagine that he was in the situation and then write down his first reaction. Self-correction was coded for responses such as "Find out what I did wrong" or "Find out my mistake," and occurred for 21 per cent of *S*s in Sample C.

Failure-avoidant interpretation. The avoidant interpretation group gave significantly more ($< .01$) denial responses than the nonavoidant interpretation group, but did not differ in aggression. Extreme groups were selected from Sample E to attempt a replication

Table 2

Specific Response Comparisons of Extreme Groups

Extreme Group	N	Incidence of Denial on SCM	Incidence of Aggression on SCM	Incidence of Self-correction on First-reaction
Failure-avoidant interpretation	18	.39	.11	Not available
Nonavoidant interpretation	30	0	.13	Not available
Corrected χ^2		9.59**	—	
High anxiety	11	.18	.72	Not available
Low anxiety	11	.18	.19	Not available
Corrected χ^2		—	4.57*	
Failure-avoidant behavior	14	.21	.28	.07
Nonavoidant behavior	14	.21	.14	.57
Corrected χ^2		—	—	6.22*

Note.—Chi square computed by actual number of *S*s rather than by proportions.
* $p < .05$.
** $p < .01$.

of this defense-specific finding, but no difference in denial was found.

Anxiety. Although no difference in denial is noted between anxiety groups, the high anxious group gave significantly more ($<.05$) aggressive responses than the low anxious group. From 245 *S*s in Sample E with anxiety scores available, the upper and lower 13 per cent were selected, and this anxiety-aggression relationship was successfully replicated (corrected $\chi^2 = 4.22$; $df = 1$; $p = < .05$).

Failure-avoidant behavior. The behavioral groups differed only in self-correction, this response occurring significantly more often ($<.05$) in the nonavoidant behavioral group than in the avoidant group. First-reaction measures were not available in other groups for replication. It might be noted that self-correction did not occur with sufficient frequency on the SCM to provide a basis for discriminating between groups.

Discussion

We have assumed earlier (4) that assimilation and failure-avoidant interpretation represent opposing poles on a continuum of reactions to negative events. However, this assumption is called into question by the failure to replicate Hypothesis I ($p < .10$). This nonsignificant difference may have resulted partially from the fact that the groups used for replication were less extreme in their interpretation scores (0–3 vs. 9–10) than the initial extreme groups. As we have discussed elsewhere (4, p. 19), the interpretation test was designed primarily to identify *S*s extreme in their interpretation tendencies rather than to measure a continuously distributed characteristic. Therefore, scores in the middle range (even as low as 3) are accordingly imprecise. Therefore, rather than inferring an avoidance-assimilation continuum, it seems safer to conclude that *S*s who are extremely avoidant will have a very low tendency to assimilate negative information. That is, if we use the mean assimilation score of 107 randomly selected *S*s (4.76) as an estimate of the assimilation base rate, the avoidant interpretation mean score (3.06) is significantly lower than this group base rate ($t = 2.90$, $p < .01$).

Similar to the present constructs of assimilation and avoidance, though used in a different context, are the processes of assimilation and contrast recently described by Hovland, Harvey, and Sherif (1) to account for individual variation in attitude change studies. Using communication statements as stimuli rather than the hypothetical negative events presently employed, these authors demonstrated that *S*s taking an extreme stand on a given issue have narrower "latitudes of acceptance" for statements which differ from their own on an issue. When exposed to such differing statements, these *S*s with extreme attitudes were found to reject the communication by contrasting it to their own position rather than by assimilating the communication. In a different stimulus realm, our avoidant interpretation *S*s may be regarded as "contrasting" negative information by their use of some nonassimilative maneuver, such as denial, which may represent a narrow range of acceptance.

Although the greater incidence of denial in avoidant interpretation *S*s would be congruent with the assumed underlying blame-avoidant tendency, this defense specific relationship must be regarded as highly tentative since replication was unsuccessful. It seems likely that the denial category requires more precise response referents and more careful attention to situational (or stem) specific determinants.

On the basis of the successful replications in both instances, the tendency for anxious *S*s to be "low assimilators" and, more specifically, to give more aggressive responses appears reasonably well established for the present population. That the high anxious *S*s do not give more denial responses than the low anxious *S*s is not surprising when one considers that denial would very likely prevent the admission of anxiety in the testing situation. The greater incidence of aggressive responses, but not denial, in high anxious *S*s agrees with earlier findings employing two similar instruments. In reporting correlates of the Test Anxiety Questionnaire, Sarason and Mandler (3) noted a significant correlation between test anxiety and the aggression subscale of the Child-Waterhouse Interfering Tendency Questionnaire (7) ($r = .29$; $N = 62$; $p < .05$). However, in considering the defensive subscale which from its definition—"need to justify or rationalize away one's failures" (7, p. 301)—appears to be similar to denial, the correlation to test anxiety did not approach significance.

The present results do not permit anchoring the assimilation tendency in behavior, although the assimilation scale differences between extreme behavioral groups were in the hypothesized direction. Although data were not available for replication, the greater incidence of a specific form of assimilation—self correction—in nonavoidant behavioral *S*s appears quite reasonable. That is, these *S*s (who may be regarded as "persistent") were selected on the basis of their remaining in the problem situation utilizing failure cues as information for developing new solutions, a behavior which would presumably require some process of self-modification.

Primarily because of its free response characteristic which allows the recovery of responses for more specific analysis such as is summarized in Table 2, the SCM, particularly in its highly structured present form, appears to provide a potentially valuable method for the initial exploration of problems in this area.

Summary

A scale measuring the tendency to assimilate negative information was developed, based on SCM responses of 800 white boys between the ages of 13 and 18. Assimilation scores were obtained from responses to five SCM stems structured as negative situations (e.g., "When someone criticizes me . . .") by three steps: (*a*) coding each response according to an objective system (inter-judge agreement = 93%); (*b*) transforming each coded response into an assimilation scale score (a three-point scale based on the likelihood that *S* would integrate the negative event into subsequent behavior); and (*c*) summing these five scores. Corrected test-retest reliability for the assimilation score over six months = .75.

Subgroups of *S*s representing extremes on three other measures were identified in order to test assimilation hypotheses: Taylor anxiety scale; failure-avoidant interpretation measure (tendency to avoid interpreting negative events as failure); and failure-avoidant behavior. The results were as follows:

1. *S*s high in anxiety scored significantly lower in assimilation tendency and gave significantly more aggressive responses than *S*s low in anxiety.
2. *S*s who showed nonavoidant behavior gave significantly more self-correction response than *S*s showing failure-avoidant behavior.
3. *S*s who gave avoidant interpretations scored lower in assimilation tendency than *S*s giving nonavoidant interpretations, but this difference was not successfully replicated.

These findings were considered in relationship to earlier assumptions regarding assimilation and avoidance; to other similar constructs (assimilation and contrast); and to other similar measures (Interfering Tendency Questionnaire). The potential value of the SCM for investigating problems in this area was noted.

Received March 20, 1957.

References

1. Hovland, C. I., Harvey, O. J., & Sherif, M. Assimilation and contrast effects in reactions to attitude change. *J. abnorm. soc. Psychol.*, 1957, **55**, 245–253.
2. Mandler, G., & Sarason, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.
3. Sarason, S. B., & Mandler, G. Some correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 810–817.
4. Schroder, H. M., & Hunt, D. E. Failure-avoidance in situational interpretation and problem solving. *Psychol. Monogr.*, 1957, **71**, No. 3 (Whole No. 432).
5. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
6. Thorndike, R. L., & Gallup, G. H. Verbal intelligence of the American adult. *J. gen. Psychol.*, 1944, **30**, 75–85.
7. Waterhouse, I. K., & Child, I. L. Frustration and the quality of performance: III. An experimental study. *J. Pers.*, 1953, **21**, 293–311.

Journal of Consulting Psychology
Vol. 22, No. 1, 1958

Q-Sort Correlation: Stability and Random Choice of Statements

Arnold H. Hilden
VA Regional Office, St. Louis and Washington University

The problem of the reliability of correlations obtained on Q sorts does not appear to have been systematically investigated. The very nature of Q sorts would seem to preclude the possibility of determining reliability through the usual approaches of alternate forms, matched items, and split halves. In a typical Q-sort experiment, an individual indicates the degree to which each item of a group describes or corresponds to the object or value in question. For example, in a self sort, he would sort statements according to how he felt they described him; in an ideal sort, he would sort them as he would like to be. Since there is no advance knowledge of his conceptions of himself, it would be impossible to construct alternate forms, to match statements, or to split a group of statements into halves of equal significance for each individual for a given frame of reference. The object toward which attention is being directed would differ uniquely for each person, since each is evaluating himself.

It seems an inescapable conclusion that no solution exists. Perhaps the restriction imposed by adherence to the usual methods creates the problem. Reliability is a matter of high importance. It reflects consistency of measurement and stability of performance. If we are not sure of the degree of consistency of measurement, we are faced with ambiguity in evaluating the changes which our instruments may reflect. The criterion of success as to reliability is the correlation between results obtained on two or more performances with an instrument or its comparable form. The method proposed here results in a high degree of consistency among 20 alternate forms chosen entirely on a random basis, with no prior evaluation of a single item of any of the 20 forms. The author has found no evidence in the literature of the use of this approach as a means of achieving consistency of measurement or stability of performance. Since the method proposed is not one of those usually employed, the question might be raised as to whether this is a measure of reliability. The question might also be raised as to whether the concept of reliability should be broadened to include the method used here.

A Universe of Personal Concepts

Any method based on matching or weighting items offers a most serious difficulty. Predetermination of values of individual items for Q sort does not seem feasible. A different approach seems indicated, one which could result in comparability of results by virtue of the method by which *groups as a whole* are selected. One possible approach would be based on sampling theory. Unbiased samples of statements from a clearly defined universe should yield results which are comparable between the groups so selected. The question, then, is the definition of the universe. The number of universes from which a selection could be made is incalculable. The first consideration would appear to be the broad nature of the universe, that is, would it involve any particular dimensions such as personality traits, or concepts of accepted clinical importance, or the formulations of some theoretical approach to personality? Whether or not there is merit in the avoidance of particular dimensions characterizing a universe of statements depends in part upon the use for which the universe is intended. Perhaps of greater importance is the difficulty of defining any universe with specific dimensions so that all possible members of it may be accounted for and identified. It is not implied that a universe can be established without any dimensions. At least one is necessary in order to establish a boundary within which all members would lie. The dimension chosen was that of *difficulty*. The universe consists of statements based on words not above the sixth-grade level. It would seem unwise to confront the majority of the population, particularly the psychological deviates, with statements employing words beyond the sixth-grade level. The use of easy words is not merely a practical consideration; it is essential, to reduce to a minimum the distortions which might affect results through lack of understanding. No matter how technically correct a formulation of human reaction might be, if the ability of the subject to understand it be disregarded, the effect may vary from person to person, and to a degree difficult to estimate. To repeat, all of the words upon which the statements comprising the universe are based are at the sixth-grade level of difficulty or below.

The members of a universe of difficulty may be identified on an objective basis through the use of Thorndike and Lorge's *The Teacher's Word Book of 30,000 Words* (4). Part I of this volume contains the 19,440 words which were found to occur at least once per million words in current popular reading. The frequency of each word is indicated by a number or letter. Accepting their statement regarding translation of frequency of occurrence into grade level (4, p. xi), words with a frequency of ten or more per million may be taken as being at the sixth-grade level or below. Thus, 6,092 words are so classified: at or below the sixth-grade level of difficulty. The next step is to select the words which are to constitute the basis for the universe.

As stated above, selection did not involve particular theories of personality, nor accepted clinical terminology, nor specific variables as such. In order to be as inclusive as possible, determination of statements included in the universe was made on the following basis. Any word was accepted for the universe if the meanings given in the *Thorndike Century Senior Dictionary* (3) could readily be used to formulate a statement of human reaction, behavior, feeling, thinking, etc. At the present stage of endeavor, it does not seem best to state the basis more specifically. Such a formulation is designated a *personal concept*. Whenever a dictionary definition provided a formulation of a personal concept, it was accepted for inclusion in the universe. In some instances, a difficulty presented itself in that many meanings were given for some words; for example, 40 meanings were given for the word "set." The selection of meanings used in this dictionary was stated to be based on Lorge's *English Semantic Count* (3, p. vii), in which the frequency of each meaning was counted. It was the use of a semantic frequency count which prompted the use of this dictionary in the present method. The order of arrangement of the meanings of a word was such that common and general meanings were put before rare and special meanings and present-day meanings put before archaic meanings. Thus, in cases where several meanings were listed, a consistent and relatively objective basis was provided for selection of the meaning to be used. Where more than one definition could be used as a basis for a personal concept, the one selected was the first one in the order listed. This procedure tended to bring into the formulations comprising the universe the meanings most frequently employed and likely to be most easily understood.

As an example, the words which were indicated to occur 10 or more times per million words on Page 1, Column 1, of the *The Teacher's Word Book of 30,000 Words* (4) were: abandon, abbey, abbot, abide, ability, able, aboard, abode, abolish. Following the procedure indicated above, three of these were used: abandon, abide, able. The respective formulations made on the basis of these words were: "I yield completely to my feelings," "I stand firm on my ideas," "I am an able person." Selecting another page at random, the words indicated on Page 94, Column 1, were: improve, improvement, impulse, in. Two of these were used: improve and impulse, for the following formulations: "I seek to improve myself," and "I act on impulse."

Following this procedure throughout the entire dictionary, a total of 1575 statements were formulated. The results of the first endeavor were set aside for a period of weeks, then carefully reviewed in their entirety to obtain a fresh approach. After another period of several weeks, they were reviewed a third time. In the compilation which constitutes the *Universe of Personal Concepts* (UPC) (2), these statements were printed in serial order, each page (but the last) having 50 statements, for a total of 32 pages. The statements on each page were numbered from 01 to 50, and the pages from 01 to 32. For example, the 20th statement on Page 3 was numbered 0320, "I trust people"; and the 35th statement on Page 6 was numbered 0635, "I am a coward."

Random Sets of Personal Concepts

With a *Universe of Personal Concepts* at hand, it was a simple matter, with random decimal digit tables (5), to obtain random sets of statements from the UPC, which have been designated *Random Sets of Personal Concepts* (RSPC). By careful adherence to the principles of random selection, some degree of confidence may be placed in each RSPC as a representation of the preferred meanings in the entire dictionary, within a given level of simplicity or difficulty. Within the limits of experimental error, some degree of confidence may also be felt that results obtained on one set of RSPC will agree with results obtained on any other RSPC and with the results on the UPC. Furthermore, the experimenter is now in a position to determine the limits of error of measurement.

Random Sets of Personal Concepts were selected from the UPC through the use of a decimal digit table (5). Selecting a point at random to enter the table, and moving horizontally to the right, four digits were taken at a time. The first pair of digits indicated the page of the UPC; the second pair, the number of the statement on that page. At each drawing, therefore, each statement in the UPC had an equal opportunity to be selected. This procedure was repeated until 50 statements were drawn, the number which had been decided upon for each RSPC. Twenty RSPC were drawn. The same personal concept could appear and did appear in different RSPC. On one occasion the same statement was drawn twice in succession for the same RSPC; the second drawing was ignored. Each statement was printed on a card 1″ × 2½″, with a code number corresponding to its seria position in the RSPC, to facilitate recording

Testing Sampling Reliability

In its broader sense, reliability is not to be conceived of simply as the correlation between two instruments or the correlation between two applications of the same instrument. The concept of reliability is embraced in a most fundamental sense when there is consistency between unbiased samples from a clearly defined universe and between these samples and the universe. Seldom is the investigator in a position to define the boundaries of a universe clearly and to identify all the members of the universe so that each has an equal chance of being drawn in any selection for a sample. When that situation does exist, the experimenter is in a position to determine *parameters*, that is, relationships involving the universe, in addition to *statistics*, that is, relationships involving samples and the universe. Stated in general terms, then, the hypothesis is advanced that the correlations obtained on Q sorts with RSPC will agree with the correlations obtained on Q sorts with the UPC.

Procedure

The subjects in this experiment were four male graduate students in psychology at Washington University.[1]

The correlations involved three kinds of relationships on each RSPC: (*a*) *r*'s between self concept and ideal concept (intra-individual, S–I), (*b*) *r*'s between self sorts of each individual and every other individual (inter-individual, S–S), and (*c*) *r*'s between ideal sort of each individual and every other individual (inter-individual, I–I).

Each subject was required to do a self sort and an ideal sort on each of the 20 RSPC and on the UPC. Cards were sorted into nine piles, with categories ranging from "most like" to "most unlike." For the self sort, the cate-

[1] Indebtedness is acknowledged to B. Buzzotta, J. Nolte, D. Petrovich, and R. Soskin, for their interest and sustained effort. Among the conditions required for a valid test of the hypothesis would be one of optimal maintenance of interest and attention to such a task. In Table 1, there is no relation between the order of their names and the order of the subjects.

gories at the extremes were "most like you" and "most unlike you"; for the ideal sort, "most like you want to be" and "most unlike you want to be."

It was postulated that the degrees of relevance or pertinence for any person for any frame of reference are normally distributed for the UPC. For nine categories of distinction, the distribution of proportions of the theoretical normal curve should be as follows: .0401, .0655, .1210, .1747, .1974, .1747, .1210, .0655, and .0401. Translating these proportions for the UPC in terms of whole numbers, the 1575 statements should be sorted into the following groups: 63, 103, 191, 275, 311, 275, 191, 103, and 63. The distribution used in this experiment conformed to these proportions. Since each RSPC was randomly drawn, the distribution of degrees of relevance for each is thereby considered normal. The distribution used in sorting each RSPC was therefore determined to be: 2, 3, 6, 9, 10, 9, 6, 3, and 2 (1).

The order in which the sorts were done was as follows: (*a*) all 20 self sorts on the RSPC, (*b*) all 20 ideal sorts on the RSPC, (*c*) the self sort on the UPC, and (*d*) the ideal sort on the UPC. After completing the 20 sorts in the first series, it seemed almost certain that no one could remember the sorting of a particular RSPC at the time of the second series. The subjects were certain they had no memory of any particular previous sort. The importance of memory in an experiment such as this is unknown, but it would seem safe to say that any effect from memory would be of negligible significance. After completion of the 20 ideal sorts, the two sorts on the UPC were done in the order stated. The execution of a Q sort constituted an assignment of weighted scores by the subject, ranging from nine at the "most like" to one at the "most unlike" end of the categories.

The data obtained from all of the sorts referred to in the preceding paragraph furnished the basis for the 336 intercorrelations in Table 1. At this point it may be well to clarify in detail what was involved in the calculation of a correlation coefficient, so that the reader may know what was correlated

Table 1

Q Sort *r*'s Obtained on 20 RSPC and the UPC on Four Subjects, Based on Intra-individual Comparisons (Self–ideal) and All Inter-individual Comparisons (Self–self and Ideal–ideal)

Random Set	Subject and Sort																Random Set
	A–A	B–B	C–C	D–D	A–B	A–C	A–D	B–C	B–D	C–D	A–B	A–C	A–D	B–C	B–D	C–D	
	S–I	S–I	S–I	S–I	S–S	S–S	S–S	S–S	S–S	S–S	I–I	I–I	I–I	I–I	I–I	I–I	
00	.78	.75	.44	.92	.73	.63	.76	.52	.75	.53	.57	.74	.79	.71	.65	.80	00
01	.80	.87	.46	.76	.54	.45	.66	.49	.82	.55	.69	.80	.78	.78	.74	.80	01
02	.73	.83	.50	.81	.66	.52	.58	.64	.82	.55	.82	.71	.76	.81	.80	.80	02
03	.75	.81	.50	.91	.60	.32	.61	.56	.68	.59	.66	.70	.70	.79	.76	.83	03
04	.84	.80	.40	.90	.57	.35	.61	.51	.75	.58	.67	.73	.73	.79	.80	.84	04
05	.85	.80	.53	.77	.71	.38	.64	.57	.67	.56	.73	.84	.77	.77	.79	.87	05
06	.85	.75	.54	.89	.51	.35	.63	.61	.77	.67	.83	.79	.77	.85	.79	.89	06
07	.82	.87	.57	.91	.70	.54	.73	.67	.74	.57	.85	.77	.76	.83	.85	.84	07
08	.83	.73	.55	.83	.63	.44	.74	.54	.60	.58	.72	.80	.83	.73	.65	.84	08
09	.90	.62	.47	.75	.55	.28	.47	.43	.77	.57	.68	.35	.75	.64	.66	.84	09
10	.83	.79	.55	.70	.64	.48	.58	.60	.66	.64	.81	.80	.86	.76	.77	.73	10
11	.84	.78	.54	.85	.47	.44	.60	.46	.68	.66	.55	.73	.80	.65	.70	.82	11
12	.72	.77	.44	.81	.61	.28	.65	.55	.69	.45	.69	.66	.75	.73	.71	.77	12
13	.69	.80	.47	.70	.65	.34	.64	.55	.62	.54	.76	.69	.63	.76	.70	.81	13
14	.91	.81	.45	.83	.54	.28	.57	.49	.59	.51	.72	.70	.70	.80	.76	.86	14
15	.77	.71	.66	.80	.69	.58	.69	.67	.62	.74	.77	.67	.76	.70	.76	.79	15
16	.83	.78	.60	.77	.55	.59	.65	.48	.62	.72	.67	.72	.76	.79	.77	.83	16
17	.77	.48	.51	.79	.42	.47	.64	.41	.65	.54	.80	.77	.84	.64	.81	.77	17
18	.82	.78	.49	.71	.56	.52	.69	.61	.60	.57	.74	.78	.79	.73	.71	.83	18
19	.77	.72	.63	.80	.39	.41	.56	.56	.46	.55	.70	.69	.82	.61	.57	.79	19
Mean *r*	.81	.76	.52	.81	.59	.43	.64	.55	.68	.58	.72	.72	.77	.74	.74	.82	
SD_r	.06	.08	.07	.07	.09	.11	.07	.07	.09	.07	.08	.10	.05	.07	.07	.04	
UPC *r*	.76	.71	.54	.76	.56	.47	.58	.56	.59	.60	.72	.69	.69	.68	.71	.83	

Note.—A, B, C and D indicate the four subjects; S indicates self sort; I indicates ideal sort. For example: A–A S–I indicates A's self-ideal sort.

with what. The subjects are designated by letters A, B, C, and D, and the Q sorts by S (for self sort) and I (for ideal sort). The RSPC were numbered from 00 through 19. Consider Subject A's self sort on RSPC 00. He sorted the 50 items into nine piles, which in effect meant he had given weights or "scores" ranging from one to nine to each item. When Subject B performed a self sort on RSPC 00, he did the same. Thus, each of the 50 items in RSPC were "scored" by the two subjects. These 50 pairs of numerical values provided the basis for the *r* of .73 found at the top of Column 5, under A–B, S–S. Again, when Subject A performed an ideal sort on RSPC, he "scored" the 50 items in that set a second time; the 50 pairs of numerical values yielded the *r* of .78 found at the top of the first column in Table 1, under the heading A–A, S–I. Each subject was compared with himself in terms of self sort vs. ideal sort on each of the 20 RSPC, resulting in 80 *r*'s for the four subjects. Each subject was also compared with each of the others in terms of their self sorts, on each of the 20 RSPC, yielding 120 *r*'s in all. Finally, each subject was compared with each of the others in terms of their ideal sorts on each of the 20 RSPC, furnishing another 120 *r*'s.

In the first column in Table 1, under A–A, S–I, are listed the *r*'s obtained from Subject A on the 20 RSPC, reflecting the agreement between his self concept and his ideal concept. The mean value and *SD* of the 20 *r*'s were calculated. The mean *r*'s and *SD*s were also determined for subjects B, C, and D. The six possible intercomparisons for self sort and ideal sort were determined in the same manner. The 16 parameters were obtained from the self sorts and ideal sorts on the UPC. It was the purpose of this experiment to determine the degree of agreement between the statistical values on the RSPC and the parametric values on the UPC.

As a null hypothesis, it would be stated that no difference exists between the mean *r*'s obtained on the RSPC and the *r*'s obtained on the UPC.

Results

From Table 1 it can be seen that not one of the differences between the mean *r*'s obtained on the RSPC and the *r*'s obtained on the UPC is statistically significant. Even at a 10% level of confidence, the null hypothesis that no difference exists would not be rejected. The degree of agreement may also be expressed in terms of the correlation between the mean RSPC *r*'s and the UPC *r*'s, $r = .94$.

Discussion

This study is offered as a methodological contribution: the establishment of a universe of statements drawn up in systematic and inclusive fashion, as a supply from which statements may be drawn for use in Q-sort studies. It has been demonstrated that statistical values for the random samples are in close agreement with the parametric values for the universe. The agreement held between persons for self sorts, between persons for ideal sorts, and within persons for self-ideal sorts. The experimenter is thus in position to determine with greater confidence the stability of Q-sort correlations for a particular individual. One need not necessarily rely on results from only one set of statements. Results from one set may be compared with results from any other set of statements. Agreement between results on the independently drawn RSPC is a basis for verification. If it is desired to see whether or not change has occurred (for example, change in a patient in psychotherapy), the same or another RSPC may be used. The use of two or more RSPC yields independent sets of measures to determine agreement as to extent and direction of change, if any. Thus, increased confidence may be felt that such changes as are reflected are meaningful in nature.

With regard to possible overlap of items in the 20 RSPC and the UPC, no actual count was made. It seems reasonable to assume that at least 600 items in the UPC were not included in any RSPC.

It may be permissible to consider an analogy in the field of physical measurement. If a random sample of adults were to be measured for height and weight, the correlation between the values would agree, within the limits of experimental error, with the correlations on other random samples and with the correlation on the entire supply from which the random samples were drawn. In executing two sorts, such as self and ideal, on a given RSPC, a subject would assign pairs of values to a set of randomly drawn statements. It should be expected that the values he would assign on random sets of statements would correspond with values he would assign on the universe from which the unbiased samples were drawn.

Summary

Reliability of Q-sort correlations cannot be determined through alternate forms involving equated items. A new approach involves selecting groups of statements rather than matching items. A universe of 1575 Personal Concepts (UPC) was drawn, using every word in the dictionary suitable for formulating human reaction at the sixth-grade level or below. Using random digits, Random Sets of Personal Concepts (RSPC) were drawn: 20 sets of 50 items. A null hypothesis was proposed: self-ideal correlations on the 20 RSPC would not differ from the parametric values on the UPC. Four graduate students performed the experiment. The hypothesis was not rejected (10% level). It was concluded that results on random sets agree with results on a universe which is clearly defined, easily understood, and consistent in frame of reference.

Received April 11, 1957.

References

1. Hilden, A. H. *Manual for Q sort and random sets of personal concepts*. Webster Groves 19, Mo. (628 Clark Ave.): Author, 1954.
2. Hilden, A. H. *Universe of personal concepts*. Webster Groves 19, Mo. (628 Clark Ave.): Author, 1954.
3. Thorndike, E. L. *Thorndike century senior dictionary*. New York: Appleton-Century, 1941.
4. Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words*. New York: Bur. of Publ., Teachers Coll., Columbia Univer., 1944.
5. U. S. Interstate Commerce Commission. *Table of 105,000 random decimal digits*. Washington: Government Printing Office, 1949.

Journal of Consulting Psychology
Vol. 22, No. 1, 1958

Manifest Anxiety in Prisoners Before and After CO_2

Dell Lebo, Robert A. Toal,[1]
Richmond Professional Institute

and Harry Brick
Virginia State Penitentiary

The Taylor Manifest Anxiety Scale (MAS) (18, 19) has figured prominently in many recent examinations of the phenomena of anxiety and learning. These studies generally have yielded findings of significance and theoretical importance. Hence, the reliability and validity of the scale as a measure of manifest anxiety deserve examination.

The reliability of several versions of the MAS, under various conditions, has ranged from .68 for anxiety scores from the complete Minnesota Multiphasic Personality Inventory (MMPI) and the Biographical Inventory (19) to .95 for the original 225-item form and a form consisting only of 50 critical items (14). A review of the literature (12) has reported a figure of .96 as the upper limit of this range. Most frequently, retest reliabilities are above .80, even after a lapse of seventeen months (10). Considering the diversity of forms marshalled under the aegis of the MAS with dissimilar buffer items (6, 9, 18, 19), changes in the number of critical items (9, 11, 18, 19), substitutions in the use of particular critical items (5, 21), and modifications of the vocabulary and sentence structure of the critical items (4, 19), these figures are reassuring.

However, the picture is not so conclusive when the aspect of validity is examined. When the MAS was first constructed it was not validated against a criterion of manifest anxiety external to the scale itself. Instead, items were selected on the basis of substantial agreement by five clinicians on 65 of 200 or so MMPI items (18). Later, 50 items showing the greatest relationship to total anxiety scores became the node of the scale (17).

As Taylor has recently pointed out, "The construction of the test was not aimed at developing a clinically useful test which would diagnose anxiety, but rather was designed solely to select *S*s differing in general drive level" (20, p. 303). Hence, many studies employing the MAS have not investigated anxiety per se, but rather the effect of drive on performance. A tidal wave of MAS studies verifying hypotheses from Hullian theory has appeared. In all of these studies anxiety has been defined in a thoroughly operational manner, i.e., in terms of MAS scores only. Such a procedure is sound methodologically but limited clinically. The need for validation against external ratings of anxiety has been assuaged with a spate of articles.

It has seemed to the present writers that an ideal way of demonstrating the validity of the MAS largely has been overlooked. This method is identical with that for validating a projective test. It consists of measuring, manipulating, and remeasuring an independent variable, in this case anxiety, to determine whether observed differences in the magnitude of the variable correlate with measured differences in MAS scale scores.

Only one study seems to have approached this technique. Gallagher (9) hypothesized that students undergoing client-centered therapy would show a significant decrease in anxiety upon the termination of therapy. He used four measures of anxiety derived from the MMPI, including the MAS. All the measures showed a significant decrease from before therapy to after. He made a point of distinguishing between *anxiety* used as a theoretical motivating factor, "present by inference rather than by direct observation," and *anxiety stress* which is "a descriptive term for a syndrome of physiological and psychological tensions" (9, p. 443).

---

[1] Now at the Medical College of Virginia.

Although Gallagher was concerned with anxiety stress, the measure he found most adequate, the MAS, has been used to demarcate both types of anxiety. Hence, the present investigators would have appreciated some indication of physiological and psychological tension in the subjects of his study other than that obtained by MMPI items and the fact that the college students presented themselves for help in personal adjustment. What seems to be missing, from the point of view of a validity study, is the direct observation of anxiety stress.

It seems germane to study anxiety in a more stressful situation than college life, in a situation where the alleviation of anxiety would be a major factor and not a "subgoal" (9, p. 444) of therapy. The present investigation used as subjects persons in a life situation exposing them to severe loss, if not to danger or injury.

Problem

The purpose of the present study was an attempt to validate the MAS directly by applying it to anxious subjects in a stress situation. The anxiety of these subjects was then deliberately manipulated in that certain of them underwent carbon dioxide therapy (CO_2) to alleviate their anxiety, while others were not treated therapeutically. It was hypothesized that as those persons receiving CO_2 had their directly observable anxiety (Gallagher's anxiety stress) reduced there would be a concomitant reduction in MAS score. Subjects not receiving CO_2 would not undergo any reduction in MAS score.

Method

Subjects

All *S*s were selected from a penitentiary with virtually all kinds of inmates whose sentences ranged from a comparatively short period to life imprisonment. Inmates appearing in the daily referral to the prison psychiatrist were screened for suitability.

> Only persons psychiatrically diagnosed as suffering from an emotional disorder with predominant symptoms of manifest anxiety were eligible to become *S*s. The *S*s selected were directly observed to be more than normally apprehensive. They presented a wide range of anxiety symptoms. Chief among these were: free-floating anxiety, periodic anxiety attacks, poor concentration, frightening dreams, sleeplessness, anorexia nervosa, fearfulness, irritability, general restlessness, and tension headaches. One *S*, who suffered from tension headaches and whose MAS score was comparatively low, asked an examiner to inspect his responses and tell him if they were those of a normal man. Other *S*s presented a more labile picture. Several *S*s feared attacks of panic. One indicated this fear by saying, "I feel like my nerves are about to overtake me." Another *S* requested treatment because he was in constant dread of being attacked by other prisoners. Although this man was well liked, his fears were aggravated by a murder which occurred in the prison. One *S* was lost to the study because of an anxiety attack while participating in the investigation. While walking down a corridor smoking a cigarette he was seen by a guard. To call the prisoner's attention to his infraction of a rule, the guard called out, "Hey you!" The prisoner began shaking, went into a severe anxiety attack, and immediately had to be hospitalized. In addition to such symptomatology, in nearly all cases the inmates' prison, social, or work adjustment was seriously impaired as a result of their anxiety.

The *S*s were not, then, given the MAS and demarcated as anxious on the basis of their test scores. Instead, they were prison inmates whose anxiety already was diagnosed psychiatrically as sufficiently serious to warrant therapeutic intervention. Twenty-four highly selected inmates were divided into treatment and control groups. There were 8 white *S*s and 4 Negroes in the control group and 7 whites and 5 Negroes in the experimental group. The *S*s were randomly assigned to either the experimental or control group in nearly all cases. It was not practicable, for example, to assign a potentially violent patient to the control group. Every effort was made to avoid assigning patients to the control group merely on the basis of a less pressing need for treatment.

The groups were closely matched with regard to length of institutionalization, age, and IQ. Statistically, the means for the two samples arose by sampling from the same population.

Procedure

A pretest interview was conducted with each *S* before testing to assess the *S*'s reaction to the test situation with particular attention to their test anxiety. *S*s who reacted in a hostile or suspicious manner were eliminated from the present investigation. The *S*s receiving CO_2 were tested within the week prior to the beginning of treatment and received their second testing session within the week after treatment was concluded. The time between the testings of the control group was identical.

Both groups continued to participate fully in institutional living. Nothing was done to alter directly the stressful environment of either group. CO_2 was administered in a matter of minutes and the inmate was returned to face the stresses of prison life. Thus, the sociopsychological effects of treatment were minimized.

Psychological Tests

The pre- and posttherapy tests consisted of the critical items of the MAS, as listed by Windle (21) with the exception of MMPI item 442, and the Bender-Gestalt test (B-G) (2). Because of suggestions that the MAS may be "especially susceptible to subject's conscious or unconscious motivation to present a favorable picture of themselves" (6, p. 277), another measure, less susceptible to faking seemed desirable. A posttherapy decrease in MAS scores should reflect a more efficient approach to reality. Pascal and Suttell (16) claimed that the B-G test is a bit of reality with which *S* has to cope. The resulting performance, therefore, may be a measure of "an attitude toward reality" (16, p. 8). Basowitz *et al.* (1), supported this contention. They found that normal, anxiety-prone *S*s in a stress situation, i.e., paratrooper training, gave less efficient performances to tachistoscopically presented B-G designs than did a control group. Since it is claimed that "simulation is impossible" (2, p. 152) on the B-G, it seemed an ideal companion, especially when objectively scored, for the "susceptible" MAS.

Carbon Dioxide

Carbon dioxide seems to be an especially effective somatic therapy in the treatment of anxiety (3, 13, 15). The present investigation employed Meduna's (15) slow coma technique, with a mixture of 30% CO_2 and 70% O_2. At no time was pure CO_2 ever introduced into the breathing bag. The number of treatments received by the experimental group ranged from 14 to 22 with a mean of 19.2. Treatments were discontinued upon manifest amelioration of anxiety, i.e., the *S*s presenting complaints were no longer directly observable and psychiatric diagnosis indicated that therapy could be discontinued. The control group received no treatment.

In his review of the literature on CO_2, Frank (8) expounded two methods of control in CO_2 experiments: (*a*) an untreated control group and (*b*) a treated control group to be given either an N_2O–O_2 mixture or O_2 "spiked" with CO_2. The second suggestion is neither as logical nor as simple as it first appears; for CO_2 is more frequently reacted to as a noxious stimulus than is N_2O. A "placebo" of N_2O could be easily distinguished from a treatment of CO_2. Spiking with CO_2 may also be unsound, for the central nervous system is highly selective in its takeup of CO_2. Consequently, a much higher concentration of CO_2 may exist in the blood than the small amount used to spike the gas would suggest. Such a concentration could not be regarded as a placebo.

Because of such considerations, the control group was untreated but continued to live in the same stressful environment as the treatment group.

Statistical Treatment

The data of the present experiment, coming as they did from a sample selected for symptoms of manifest anxiety, could be expected to be markedly skewed. The scoring of the B-G also distributed in a moderately skewed manner in the population on which the test was standardized (16). Hence, Festinger's (7) distribution-free method of analysis was employed.

Results and Discussion

Median scores for the experimental and control groups of the MAS and the B-G are shown in Table 1. The Festinger (7) rank

Table 1

Median Scores of Experimental and Control Groups on the MAS and the B-G

Test	Experimental Group		Control Group	
	Pretest	Posttest	Pretest	Posttest
MAS	28.5	19.0	22.0	22.5
B-G	30.5	22.0	20.5	33.0

difference comparisons of the two groups are listed in Table 2.

As can be seen from Table 1, both the MAS and B-G scores for the experimental group decreased following CO_2. MAS values for the control group remained approximately the same, suggesting the environment they originally found to be stressful had remained relatively unchanged. The B-G scores of this group, however, markedly increased. This difference in scores is interesting, but because of the limited number of *S*s the present writers will not indulge in public *post hoc* speculation.

Table 2 reveals that the MAS scores decreased in a statistically significant manner for the group whose anxiety was treated by CO_2. There was no such decrease for the untreated anxious group. These people, presumably, continued to operate in a stressful milieu. The hypothesized reduction in MAS score was realized.

The MAS may be said to reflect differences in the magnitude of the variable of manifest anxiety. Hence, it was found to be a valid measure of manifest anxiety.

It should be remembered that the B-G test was included to check any tendency on the part of the experimental group to fake MAS responses. The improvement of the treatment group was significant also in their B-G scores. As MAS scores decreased B-G performance improved. This improvement in coping with the reality of the B-G test suggested that the MAS was not faked by the experimental group.

Table 2

Festinger Comparisons of Experimental and Control Groups on the MAS and the B-G

Test	Mean Rank		
	Experimental Group	Control Group	*d*
MAS	9.08	15.92	3.42*
B-G	9.42	15.58	3.08*

* Significant at the .05 level.

An association implicit in the present study is that the observation of manifest anxiety has a relationship to MAS scores. Several studies specifically relating MAS scores and observational data have appeared. In none of these studies has the relationship between observed anxiety, observed diminution of anxiety, and MAS scores been investigated. Certainly the high relevance of this relationship, quantified and made explicit, earmarks it as a desideratum for future research.

Summary

Distribution-free comparisons of a group of 24 manifestly anxious prisoners were made. Half of the men received CO_2 and half were given no treatment. Reasons for the absence of a placebo control were advanced. A statistically significant improvement in the performance of the experimental group on the MAS was obtained. This improvement was also seen in a check test, the B-G. The results were interpreted as indicating the validity of the Taylor MAS as a measure of manifest anxiety.

Received March 18, 1957.

References

1. Basowitz, H., Persky, H., Korchin, S. J., & Grinker, R. R. *Anxiety and stress.* New York: McGraw-Hill, 1955.
2. Bender, Lauretta. *A visual motor gestalt test and its clinical use.* New York: Amer. Orthopsychiat. Ass., 1953.
3. Brick, H. Carbon dioxide therapy in private practice and in the Virginia State Penitentiary. *Tri-State Med. J.*, 1956, **3**, 14–17.
4. Buss, A. H. A follow-up item analysis of the Taylor anxiety scale. *J. clin. Psychol.*, 1955, **11**, 409–410.
5. Cook, W. E. A study of anxiety in the Minnesota Multiphasic Personality Inventory. *Va. J. Sci.*, 1953, **4**, 28.
6. Davids, A. Relations among several objective measures of anxiety under different conditions of motivation. *J. consult. Psychol.*, 1955, **19**, 275–279.

7. Festinger, L. The significance of the difference between means without reference to the frequency distribution function. *Psychometrika,* 1946, 11, 97–105.
8. Frank, J. A. A critical evaluation of carbon dioxide oxygen inhalation therapy in mental disorders. *Amer. J. Psychiat.,* 1953, 110, 93–103.
9. Gallagher, J. J. Manifest anxiety changes concomitant with client-centered therapy. *J. consult. Psychol.,* 1953, 17, 443–446.
10. Hedlund, J. L., Farber, I. E., & Bechtoldt, H. P. Normative characteristics of the Manifest Anxiety Scale. Paper read at Midwest. Psychol. Ass., Chicago, April, 1951.
11. Holtzmann, W. H., Calvin, A. D., & Bitterman, M. E. New evidence for the validity of Taylor's Manifest Anxiety Scale. *J. abnorm. Soc. Psychol.,* 1952, 47, 853–854.
12. Kendall, E. The validity of Taylor's Manifest Anxiety Scale. *J. consult. Psychol.,* 1954, 18, 429–432.
13. LaVerne, A. A., & Herman, M. An evaluation of carbon dioxide therapy. *Amer. J. Psychiat.,* 1955, 112, 107–113.
14. Matarazzo, J. D., Guze, S. B., & Matarazzo, Ruth G. An approach to the validity of the Taylor anxiety scale: scores of medical and psychiatric patients. *J. abnorm. soc. Psychol.,* 1955, 51, 276–280.
15. Meduna, L. J. *Carbon dioxide therapy.* Springfield, Ill.: Thomas, 1950.
16. Pascal, G. R., & Suttell, Barbara, J. *The Bender-Gestalt test.* New York: Grune & Stratton, 1951.
17. Spence, K. W., & Taylor, Janet A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.,* 1951, 42, 183–188.
18. Taylor, Janet A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.,* 1951, 41, 81–92.
19. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.,* 1953, 48, 285–290.
20. Taylor, Janet A. Drive theory and manifest anxiety. *Psychol. Bull.,* 1956, 53, 303–320.
21. Windle, C. The relationships among five MMPI "anxiety" indices. *J. consult. Psychol.,* 1955, 19, 61–63.

Journal of Consulting Psychology
Vol. 22, No. 1, 1958

Bender-Gestalt Test Correlates of Emotional Depression[1]

John E. Tucker and Mimi J. Spielberg
Mental Hygiene Clinic, VA Hospital, Albany, N. Y.

The purpose of this study was to discriminate between depressed and nondepressed clinical patients by means of the Bender-Gestalt given in diagnostic test batteries.

Thirty-six male *Ss* were divided into two groups, depressed and nondepressed. The differences in age and intelligence between the two groups were not statistically significant. The sample included diagnoses of psychosis, psychoneurosis, and character disorder, and excluded diagnoses of organic brain damage.

Trained raters determined the presence or absence of depression by a standardized evaluation of the *Ss*' VA clinical records and psychological test materials, excluding the Bender. Acceptable agreement between raters was demonstrated in the differentiation of the two groups. A 27-item check list was devised for use as a guide and summary in judging depression.

The Bender scoring system used was that of Pascal and Suttell, a 107-item check list. The identity of the patient was concealed.

Comparisons were made between the depressed and nondepressed *Ss* on 20 separate Bender test item scores. Not all the items selected for comparison were equally relevant clinically. Of the 20, only two items, tremor and distortion, were found to be significant at the .05 level. No items were significant at the .01 level.

It can be concluded that *Ss* sometimes express depression through a tremor on the Bender-Gestalt. Tremor alone is not sufficient to diagnose depression, but it perhaps can be used to warn a clinician of that possibility. Tremor was found primarily among the psychoneurotic *Ss*.

The greater distortion was found in the depressed group, and the *Ss* who distorted the designs were primarily psychoneurotics. No distortion was found in the nondepressed group. This finding seems to contradict the hypothesis of pedantic accuracy usually associated with depressed behavior. Since this result does not appear to be clinically meaningful, it may be an artifact of the sample. Other items usually thought to be related to depression, such as compression of drawing space, were nondiscriminating. The Bender test, as scored by Pascal's system, does not seem to offer clinical usefulness in detecting emotional depression in typical cases in our clinic.

Comparisons were made of average initial reaction time and average total response time to each card. It is interesting to note that the tendencies were not in the expected direction. The nondepressed *Ss* were slower in the Bender response times than the depressed patients.

Brief Report.
Received October 25, 1957.

---

[1] An extended report of this study may be obtained without charge from John E. Tucker, Mental Hygiene Clinic, VA Hospital, Albany, N. Y., or for a fee from the American Documentation Institute. Order Document No. 5430, remitting $1.25 for microfilm or $1.25 for photocopies.

Journal of Consulting Psychology
Vol. 22, No. 1, 1958

Assaultiveness and Two Types of Rorschach Color Responses[1]

Robert Sommer and Dorothy Twente Sommer
The Saskatchewan Hospital, Weyburn

Much of the thinking about the relationship between Rorschach responses and other behaviors is based on the concept that the formal aspects of the responses (determinants) should have correlates in other behaviors of the person (introversion, anxiety, etc.). However, present-day thinking tends to regard information as to structural aspects of personality alone (e.g., defense mechanisms) as insufficient for understanding the behavior of the individual. Past thinking was considered unsatisfactory in that it looked for relationships between energy, affects, or need and overt behavior without taking into account the channels the organism had developed for handling this energy. It was also felt that information as to these structures alone left one with a bare shell and still unable to predict the actual behavior of the person. What was needed was information regarding both the contents of fantasy *and* the person's techniques for handling his affects.

At least two approaches to Rorschach research are unsatisfactory: (*a*) to relate determinant scores to behavior without regard to the content of the responses;[2] (*b*) to relate content categories (e.g., destructive, oral expulsive, etc.) to behavior without regard to the formal aspects of the response (determinant score).

An attempt is made in the present study to base predictions on a combination of formal and content elements. The type of behavior selected for study is assaultive or explosive behavior where the Rorschach literature is both considerable (3, 4, 6, 7) and often contradictory. There have been several attempts to design experiments that would assess the merits of the hypotheses proposed, but the results are not easy to understand. With regard to structural correlates of explosive behavior, there are those who believe that either very many color responses or no color responses are indicative of it. With regard to the content of the responses, predictions can be both very specific ("tomahawk" to the white space on Card VII indicates aggressive acting out) or very general (many animal responses indicates stereotyped thinking). In the prediction of aggressive behavior from the Rorschach, there are two distinct groups of clinicians: (*a*) Some believe aggressive contents indicate that the person is able to express his hostile impulses in a socially acceptable way and hence has no need to act out. Phillips and Smith (5), for example, speak of blood responses as contraindicative of destructive acting out. (*b*) Others believe that the use of hostile content indicates a preoccupation with such thoughts. Depending upon the individual's impulse-control system, such behavior should be expected if the sample of the individual's life-history is adequate. Few tests have been based on the first approach except that of Stone (7) which met with essentially negative results. Those of Pitluck (6) and Finney (1) may be taken to support the latter view, although in the former case the results were not clear-cut.

---

[1] The study was supported by a grant from the State of Louisiana hospital research fund. Without the active cooperation of the staffs of Gulfport VA Hospital (especially Herdis A. Deabler) and of Southeast Louisiana Hospital (especially Joseph G. Dawson), the study could not have been undertaken.

[2] With some determinants there is a confounding of content and structure, as in *M* where the use of a human figure is usually required. Since this is a common but not necessary condition for *M*, it makes the meaning of any correlation between *M* and other factors quite ambiguous.

The present problem is to determine the relationship between explosive behavior and two types of Rorschach color responses: (*a*) aggressive and explosive responses (volcano, fire, blood from a wound, etc.); (*b*) nonaggressive responses (bouquet, ice-cream, orchid, etc.). One essential condition is that the two groups should not differ in *Sum C;* the only difference should be in the content of the color responses. Our prediction, based on the work of Stone, Pitluck, and Finney, is that the "aggressive content" group will show more explosive and assaultive behavior than the "nonaggressive content" group. It was also decided to distinguish between physical explosiveness and assaultiveness (defined as explosive physical attacks on persons, objects, or animals in the environment) and verbal explosiveness and assaultiveness (explosive episodes where torrents of abuse are directed toward the outside world). Both definitions stipulate that the behaviors be characterized by an explosive quality in addition to being assaults upon the environment.

An ancillary hypothesis was that "aggressive movement responses" (e.g., fighting, kicking, etc.), when accompanying aggressive color responses, would improve the prediction of aggressive acting-out behavior. However, the chief emphasis of the study was on the differences in behavior between *S*s who gave "aggressive color responses" and those who gave "nonaggressive color responses."

Procedure

Our preference for criteria of assaultiveness was for meaningful acts by the individual in his own particular world. This type of material is most easily found by interviewing or in good case histories if they happen to be available. Yet if these are gathered for purposes other than for one's experiment (e.g., hospital records), one must always be aware that the interviewer may not have been interested in the behavior that one is studying. If the behavior studied is of considerable importance in the individual's relations with others, as is the case with assaultive behavior, the chances of the interviewer's overlooking it are minimal, although the possibility must still be considered.

Hence, our criteria of assaultiveness are incidents of such behavior in case histories gathered routinely by the social service department of a VA hospital. The procedure limited our sample to male psychiatric patients which is regrettable in terms of generalizing from the results but advantageous if aggressive acting-out is more prevalent in such a group.

Essentially, the method embodied an *ex post facto* design in which two investigators worked independently. The senior author spent several weeks going through the files of the psychology department of this hospital and looked over approximately 200 protocols (none of which he had seen before). He wrote down the names of those patients who had given explosive color responses (objectively defined as responses of volcanoes, explosions, fire, etc.) and those who had given exclusively nonaggressive color responses (orchids, bouquets, sherbert, etc.) and also whether or not the *M* responses could be classified as aggressive or not (objectively defined as actions of fighting, kicking, etc. in a way similar to Finney's category of "active destruction" in his Palo Alto scale). If a person with aggressive color responses also had several nonaggressive color responses, he was still included in the aggressive color category. If a subject had even six or seven nonaggressive color responses and one that was partially aggressive, he was *not* included in the nonaggressive color group.

The list of approximately 65 names, arranged in alphabetical order and with all other information removed, was given to the second investigator who was instructed to inspect the case history of the patient and rate the prehospital behavior along a seven-point scale for (*a*) physical assaultiveness and (*b*) verbal assaultiveness. If there was insufficient information for rating along a dimension, a zero could be used. Some attrition was expected in our sample as a few case histories had been sent to other installations or could not otherwise be located.

Results

The final data consisted of the ratings of the case histories of 26 aggressive color *S*s

Table 1

Ratings on Explosiveness and Assaultiveness

Measure and Group	N^a	Mean	S^2	t	p
A. Physical explosiveness					
Aggressive color *S*s	22	3.41	3.14	1.69	.05
Nonaggressive color *S*s	26	2.54	2.93		
B. Verbal explosiveness					
Aggressive color *S*s	23	3.78	2.45	1.61	N.S.
Nonaggressive color *S*s	28	3.07	2.22		
C. Physical and verbal explosiveness					
Aggressive color *S*s	22	7.14	10.79	1.67	.05
Nonaggressive color *S*s	25	5.52	10.33		

[a] The case histories of some *S*s did not contain sufficient information for rating along all dimensions. Hence, the figures in the cells may vary slightly depending on the dimension compared.

and 31 nonaggressive color *S*s. The ratings on both physical and verbal assaultiveness ranged over the full scale (from 1, very low, to 7, very high). The mean rating of physical assaultiveness for the total population was 2.98 while for verbal assaultiveness it was 3.39. The mean *Sum C* for the aggressive color group was 3.83, while for the nonaggressive color group it was 3.60. Neither of these differences are significant.

The data which test the hypothesis that aggressive color *S*s exceed the nonaggressive color *S*s in ratings on assaultiveness are presented in Table 1.

For physical assaultiveness, the aggressive color *S*s significantly exceed the nonaggressive color *S*s. The results on verbal assaultiveness are suggestive but fall slightly short of significance. If the ratings for the two types of assaultiveness are pooled to form a rating on general assaultiveness, the difference between the groups is significant at the .05 level. The biserial correlation between the type of color response given by the *S* and the pooled ratings of aggressiveness is .35 ($p < .01$).

Table 2

Percentage of Cases Predicted

Assaultiveness Rating	Rorschach Rating: Aggressive Color *S*s	Rorschach Rating: Nonaggressive Color *S*s
Above median	59	40
Below median	41	60

Table 2 presents the percentage of correct and incorrect predictions when the assaultiveness ratings are divided at the median. It shows that although the results are in the expected direction, it would be hazardous to depend on this variable alone in making predictions about the potentiality for assaultive behavior in individual cases.

Some interesting results appear when the protocols are divided on the basis of *S*s who gave aggressive, nonaggressive, and no movement responses. An overall rating of the movement responses had been made by the senior author at the time of classification of color responses. If there was doubt as to whether the tone was aggressive or nonaggressive, the cases were placed in a separate category and not included in the comparisons in Table 3.

Although the number of cases is necessarily smaller than in the preceding table, the differences are even more marked when based on type of *C* response *and* type of *M* response. In terms of the predictive value of a combination of movements to color responses, the biserial r for a comparison of the "aggressive color and aggressive or no movement *S*s" with the "nonaggressive color and nonaggressive or no movement *S*s" for their hostility ratings is .55 ($p < .01$). There are not enough cases where aggressive color *S*s gave clearly aggressive *M* responses to use either of these categories in the comparisons.

Table 3

Ratings on Explosiveness and Assaultiveness with *S*s Classified by Type of *C* and *M* Responses

Measure and Group	*N*	Mean	*t*	*p*
A. Physical explosiveness				
Aggressive *C* + Aggressive *M*	8	4.00	1.80	.05
Nonaggressive *C* + Nonaggressive *M*	12	2.67		
Aggressive *C* with no *M*	6	4.00	1.38	N.S.
Nonaggressive *C* + Nonaggressive *M*	12	2.67		
Aggressive *C* + Aggressive *M*	8	4.00	3.33	.01
Nonaggressive *C* with no *M*	8	1.75		
Aggressive *C* with no *M*	6	4.00	2.46	.05
Nonaggressive *C* with no *M*	8	1.75		
B. Verbal explosiveness				
Aggressive *C* + Aggressive *M*	8	4.13	1.44	N.S.
Nonaggressive *C* + Nonaggressive *M*	13	3.31		
Aggressive *C* with no *M*	6	4.17	.86	N.S.
Nonaggressive *C* with Nonaggressive *M*	13	3.31		
Aggressive *C* + Aggressive *M*	8	4.13	3.57	.01
Nonaggressive *C* with no *M*	9	2.22		
Aggressive *C* with no *M*	6	4.17	1.99	.05
Nonaggressive *C* with no *M*	9	2.22		

Although the two groups of *S*s had approximately the same *Sum C* and *M* totals, the *Sum C* of the aggressive color *S*s contained more pure *C* responses than that of the nonaggressive color group, which in turn exceeded the former group in number of *FC* responses. The *CF* responses were approximately the same for the two groups. This result was hardly unexpected as most of the aggressive color responses could be scored *C* or *CF*, while "flowers" or "grasshoppers" generally rate scores of *FC* or at most, *CF*. It seemed in order to see if a division by determinant alone (*FC* versus *CF* + *C*) would prove to be of more predictive utility with the ratings on assaultiveness than the use of content regardless of form-color balance. Finney found that his groups of aggressive and nonaggressive *S*s differed with regard to their *CF* + *C*:*FC* ratios.

Of our sample, 39 *S*s had *CF* + *C* exceeding *FC*, while 11 had *FC* equalling or greater than *CF* + *C*. Analysis of these data showed that there was not a significant difference between the two groups on either physical or verbal assaultiveness.

Discussion

There were two theories about aggressive Rorschach color responses outlined at the beginning of this paper. The first was that such responses constituted a draining off of hostile impulses and we should find less acting-out in such *S*s. This would postulate a concept of a level of unreality where the person can solve his problems without fear of retribution from a hostile environment. If he is able to do this, he will have no need to act-out in a socially unacceptable fashion. There is an implicit assumption here of a specific amount of psychic energy that the person can release either on a fantasy level or on a behavioral level or on both.

Two groups of empirical studies support this view. The first would be instances where tension released in overt behavior reduces the tension in the fantasy level. A case in point would be the Zeigarnik effect experiments where the act of completing certain tasks will reduce the psychic tension as measured by memorial effects. There are also the studies carried out at Clark University where physical restraint increased the number of *M* responses given on the Rorschach.

Conversely, there are those who believe that the use of a content area indicates a concern or even preoccupation with a given topic. Hence, we should certainly expect behavioral correlates if we look hard enough. This view implies that the Rorschach gives a sample of behavior in a particular situation (involving an interpersonal relationship, a cognitive task, and a subjective experience on the patient's part of being evaluated). We should logically expect to find other behavioral correlates of Rorschach behavior. Loose or uncontrolled responses to (hypothesized) affect-arousing stimuli should parallel analogous behaviors in similar situations. The patient who "explodes" during the Rorschach (or phrased less strongly, the patient who loses the ability to control his response to color) would be the person we might expect to "explode" in other situations.

Our results support this second view. Aggressive responses on the Rorschach are *not* negatively corrected with aggressive acting-out, they tend to be positively correlated with it. This has implications both for making predictions of such behavior from the Rorshach and for our understanding of the relationship between fantasy [3] and behavior. The "catharsis" view of test responses is not uncommon among clinicians, especially in the matter of blatant sex responses. The idea is often proposed that the person who reports several penises and vaginas will be a mature person who is capable of expressing his impulses in an appropriate manner and hence less likely to indulge in antisocial sexual behavior. Our hypothesis would be that this is an incorrect supposition. Research should show that individuals committing rape or indulging in self-destructive promiscuous behaviors will produce significantly more blatant sex responses than patients hospitalized for other reasons. The hypothesis should hold even for veiled or suggestive test responses. One can recall the Levine, Chein, and Murphy (2) study in which responses of objects instrumental in eating (in addition to actual food responses) increased with food deprivation. Our hypothesis would be that except for rare instances, such as the autism of the schizophrenic, can fantasy serve as a *contraindication* of analogous overt behavior. Even the schizophrenic girl preoccupied with sexual fantasies will display seductive and exhibitionistic behaviors on many occasions. It is our view that evidence for the "catharsis theory of test responses" is based largely on inadequate samples of the person's behavior. It may be true that the girl with very strong sex fantasies will act prim and prudish when we interview her. Yet examine her behavior over the past year, or five years, or when she "let herself" have a few drinks, then we will find marked resemblances between fantasy life and other behaviors. Fortunately, in using case histories, we were able to view a sizable portion of the individual's life. In studies involving artificial stimulus situations or small isolated samples of behavior, we should not expect to find high correlations between fantasy contents and behavior.

Summary

The present study aimed at assessing the relationship between assaultive behavior and two types of color responses, aggressive and nonaggressive. The prediction was made that *S*s giving aggressive color responses should show more assaultive behavior than the nonaggressive color *S*s.

One investigator looked through approximately 200 Rorschach protocols of male patients in a VA hospital and listed the names of *S*s giving aggressive color responses and those giving nonaggressive color responses. The second investigator was given the list (without any information as to the category of the *S*) and was asked to rate the case history of the *S* as to incidents of verbal and physical assaultiveness. The hypothesis was confirmed. The trends were especially clear in cases where *S* had given both aggressive color *and* aggressive movement responses.

---

[3] "Fantasy" as used here refers to an unknowable construct, literally "the person's inner world," the contents of which can only be inferred from verbalizations, parapraxes, etc.

Although the results were of theoretical relevance in assessing the relationship between fantasy and overt behavior, the correlations were not of sufficient magnitude to permit their use in individual prediction of assaultive behavior.

Received March 25, 1957.

References

1. Finney, B. Rorschach test correlates of assaultive behavior. *J. proj. Tech.*, 1955, **19**, 6–17.
2. Levine, R., Chein, I., & Murphy, G. The relation of the intensity of a need to the amount of perceptual distortion: A preliminary report. *J. Psychol.*, 1942, **13**, 283–293.
3. Lindner, R. M. The Rorschach test and the diagnosis of psychopathic personality. *J. crim. Psychopath.*, 1943, **5**, 69–93.
4. Lubar, G. H. Rorschach content analysis. *J. clin. Psychopath.*, 1948, **9**, 146–152.
5. Phillips, L., & Smith, J. G. *Rorschach interpretation: Advanced technique.* New York: Grune & Stratton, 1953.
6. Pitluck, Patricia. The relation between aggressive phantasy and overt behavior. Unpublished doctoral dissertation, Yale Univer., 1950.
7. Stone, H. The relationship of hostile-aggressive behavior to aggressive content on the Rorschach and Thematic Apperception Test. Unpublished doctoral dissertation, U.C.L.A., 1953.

Journal of Consulting Psychology
Vol. 22, No. 1, 1958

A Normative Note on Sentence Completion Cross-Sex Identification Responses

Francis W. King
Dartmouth College

The interpretation of projective responses may depend heavily for years upon such real, if elusive, factors as clinical experience and judgment. Quantitative data, however, can make more explicit and more communicable some aspects of clinical experience; furthermore, it can correct individual bias in interpretation.

Normative data establish a more objective frame of reference and should serve in some instances to taper the more extravagant inferences. The meaning of cross-sex identification responses is a case in point. Eron (1) has reported that 30.7% of the nonhospitalized males in his normative TAT study immediately misidentified the sex of the figure in Card 3BM; an additional 29.4% evidenced uncertainty as to the figure's sex, with a subsequent 18.7% misidentifying and 10.7% correctly identifying. On the other hand, Mainord (3) found that 94.7% of her male subjects in a figure-drawing study drew the self-sex (male figure) first. With these normative differences, it would seem that the clinician who makes comparable interpretations from such cross-sex identifications does so at his peril.

In view of the near absence of normative data on sentence completion responses, the writer undertook the following investigation. A sentence completion form was administered to an incoming freshman class prior to the beginning of classes. Using a sample of 212 male students, who subsequently took the elementary psychology course, the writer analyzed the results of sentence stems that yielded data on identification. The following three items produced cross-sex identification responses.

23. When he found the trunk with all the clothes in it, just for fun he dressed up as ——
39. In the college play he took the part of ——
43. He hoped that some day he would be just like ——

The results in Table 1 indicate that cross-sex identification responses are to some extent a function of the stimulus properties of the sentence stems, or, if one prefers, a function of the differing levels of personality tapped by such stems. Thus, Item 43, a rather "obvious" item, produces cross-sex identifications in only 1% of the population studied; in most cases it evokes a variety of fairly "public" expressions of ego ideal or desires readily accessible to awareness. Item 23, in contrast, produces a rather large number (20%) of cross-sex identifications, as well as some primitive and regressive responses revealing identification with animals and children. Different sentence stems evoke responses from differing levels of personality organization; this clinical observation is borne out by the research of Hanfmann and Getzels (2). The whole area is one which should be explored systematically and thoroughly if the greatest usefulness is to be derived from this easily administered and productive technique. Of the 43 subjects who responded with feminine identifications on Item 23, only three gave female responses to Item 39 and only one to Item 43; an additional five gave "residual" responses in Item 39. This residual category contains a miscellany of responses that are nonspecific with respect to sex; some of these certainly reflect other than healthy processes, e.g., "death," "a nobody," "dope." No subject responded with feminine identifications in both Item 39 and Item 43. It might be mentioned in pass-

Table 1

Summary of Responses by Categories in Sentence Stems Producing Cross-Sex Identifications ($N = 212$ college males)

Identification Response	N	%
Item 23		
Male identification	143	67
Female identification	43	20
Animal identification	4	2
Child identification	2	1
Residual miscellaneous responses	20	9
Item 39		
Male identification	177	83
Female identification	6	3
Animal identification	4	2
Child identification	1	0.5
Residual miscellaneous responses	24	11
Item 43		
Father identification	112	53
Female identification	2	1
Identification with other male relative	20	9
Identifiable male figure (real or fictional)	16	8
Ideal or hero	11	5
Occupational responses	9	4
Residual male figure	27	13
Responses emphasizing individual uniqueness	8	4
Residual miscellaneous responses	7	3

ing that two subjects replied with animal identifications in both Items 39 and 43.

In addition to evoking responses which lend themselves to classification with some ease (especially Items 23 and 39), these stems produce idiosyncratic responses which are clinically rich and serve well as bases for hypotheses concerning the individual's personality organization with reference to unconscious needs, ego ideal, fantasied self-image, and the like. The wide variety of individual replies to such sentence stems insures that the intuitive and imaginative clinician will have ample opportunity to exercise his skills and also makes it obvious that the classificatory tone of this note cannot be suggestive of a substitute method of analysis.

The problem of making valid interpretations of an individual's responses to projective techniques will never be resolved by a completely normative approach. Those who prize the clinical mystery need not fear that such techniques will be reduced to a clerical operation. However, the results presented here should serve to caution the interpreter to take due cognizance of the variability in projective content evoked by different stimulus materials—in this case different sentence stems. Furthermore, here are provided some beginning normative data on three clinically useful items.

Received March 25, 1957.

References

1. Eron, L. D. A normative study of the Thematic Apperception Test. *Psychol. Monogr.*, 1950, **64**, No. 9 (Whole No. 315).
2. Hanfmann, Eugenia, & Getzels, J. W. Studies of the Sentence Completion Test. *J. proj. Tech.*, 1953, **17**, 280–294.
3. Mainord, Florence R. A note on the use of figure drawings in the diagnosis of sexual inversion. *J. clin. Psychol.*, 1953, **9**, 188–189.

Journal of Consulting Psychology
Vol. 22, No. 1, 1958

Some Determinants of the Perception of Hostility

Bernard I. Murstein[1]

University of Texas, M. D. Anderson Hospital and Tumor Institute

The relationship between the responses obtained through the use of projective techniques such as the Rorschach and the overt behavior of persons has been one of the persistent problems confronting the clinical psychologist. Gluck, in studies with the TAT and Rorschach (3, 4), found no correlation between overt hostile behavior and overt and covert signs of hostility on the Rorschach and TAT. Cattell (1) has stated a theory holding for a proportionality of overt and covert personality deviation. Murstein (5), defining projection as "the ascribing of one's own motivations, feelings and behavior to other persons" referred to five variables which he considered to be functions of projection: (*a*) the objective characteristics of the person, (*b*) the self-concept, (*c*) the amenability of the trait investigated to measurement, (*d*) the conditions under which projection is measured, and (*e*) the instrument used to measure projection.

Findings related to each of these variables have also been reported. All of them were of some significance in either of two conditions (Rorschach, ego threat) investigated. Gluck's failure to find any relationship between the Rorschach and overt behavior may well have been a function of the failure to take cognizance of the above variables.

Granted that these variables are important in the prediction of the operation of the mechanism of projection, the question arises as to whether there are any significant relationships between the tendency to project hostility on the Rorschach and the tendency to perceive hostility as the result of an ego-threatening situation. The procedure for investigating this question follows.

Procedure

The method of pooled ranks was used to obtain group and self judgments of hostility on 536 students from 23 fraternities and two dormitories at The University of Texas by means of cutting scores based on consideration of the standard error of measurement. The method has been discussed more fully elsewhere (5). Eighty men, divided equally among four experimental groups (hostile-insightful, hostile-noninsightful, friendly-insightful, and friendly noninsightful) were selected.

Each *S* of every group was given a Rorschach test without formal inquiry for determinants. Only the content, animation, and description of the perceptions was recorded, in a manner similar to that followed by Elizur in his Rorschach Content Test (2). The instructions given to each *S* were:

> I am going to show you a series of cards which are in reality inkblots. These inkblots have been chosen because they resemble many things. I want you to look at each of these cards and to tell me the first three things that you see on them. Do not hesitate to tell me anything that you see. Try to give a full description of what you see rather than stopping after you have given just the name. There are no right or wrong answers as this is not a test. There are no restrictions at all as far as what you see. It is entirely up to you. I will write down on this sheet of paper whatever you tell me that you see. Are there any questions?

After all questions had been answered, the Rorschach cards were presented in the usual manner. At the conclusion of the administration of the cards, the examiner, after carefully glancing over the *S*'s record, said, "Now I shall give you a brief interpretation of what you have seen."

Each experimental group of 20 *S*s was divided and each half matched according to the means of group rankings and self-rankings and their respective variances. Within each group, one half received the interpretation "friendly," while the other half received the interpretation "hostile." Initially, in selecting the *S*s, the experimental group to which each *S* belonged was of course known. The author detached the name from the group so that at the time of testing (five weeks later) all that was known of any *S* was whether he was to receive a "friendly" or a "hostile" interpretation. The purpose of this procedure was to forestall the possibility of bias on the part of the author which might favor one group over another.

---

[1] Now at Louisiana State University.

The following points were stated in the "friendly" report:

> 1. You cooperated very nicely in taking this test. You saw things readily because you were interested and really "put yourself" into it. 2. Your perceptions are very rich in creativity and imagination. 3. Your perceptions reveal a lot of feeling for people, and a lot of warmth; you are a friendly, cooperative person. 4. Your perceptions indicate a deep sensitivity to the needs of others. 5. You are, therefore, psychologically speaking, a mature and fairly well-adjusted person. Do you have any questions?

The following points were stressed in the "hostile" report:

> 1. You showed a lack of cooperation in taking this test. Your perceptions indicate that you were bored, disinterested, and didn't bother with the test. 2. Your perceptions are, accordingly, poor in imagination and indicate a lack of creativity. 3. You are pretty "cold" toward people and an uncooperative, hostile person. 4. Your perceptions indicate a lack of sensitivity to the needs of others. 5. You are, therefore, psychologically speaking, immature and not too well-adjusted. Do you have any questions?
>
> The most frequent question was, "Where did you get that from?" In this case, a response was chosen at random and the examiner merely stated again that this had either a "friendly" or "hostile" connotation (accordingly), at a deeper psychological level.

The examiner then presented the Examiner Rating Sheet (henceforth referred to as ERS) to the *S*, saying,

> Now, I shall ask you to fill out anonymously a rating sheet whereby *you* evaluate me. This is a part of some research that the psychology department is currently running. Do not put your name on this sheet. When you have finished the ratings, put the sheets in this envelope and seal it and drop it in this box. I shall return in about five minutes to see whether or not you are finished and ready to go on to the last part of the study.

The ERS consisted of 19 statements taken from the larger Interview Rating Scale used by The University of Texas Testing and Guidance Bureau (7). A typical statement taken from this sheet was "The interviewer is a warm, sincere individual." The *S* rated each statement on a five-point scale: five for "strongly agree," four for "agree somewhat," three for "undecided," two for "doubtful," and one for "strongly disagree." The higher the score, the more favorably the *S* rated the interviewer.

The ERS was filled out and returned, and then the examiner explained that the report which he had given the *S* was used for experimental purposes and was not meaningful. The random nature of selection of the "hostile" report was explained, and the fact that one person out of two in the total group received a "hostile" report seemed to reassure most of the *S*s. None of the *S*s seemed disturbed by the experiment. Before leaving, the *S*s were asked not to mention any aspect of the experiment to anyone for four weeks, and were also thanked for their participation. To measure the projection of hostility on the Rorschach, a new scale was devised called the Rorschach Hostility Scale (RHS). This scale has a range of from one to seven points for any given response on the Rorschach. It is essentially a two dimensional scale, one dimension being an overt-covert one, while the other is a lower-phyla–higher-phyla range. In other words, a mangled human would be scored higher with regard to hostility than a mangled insect. This scale has been shown to be quite reliable when scored by trained scorers, with an average reliability of .96 (4). The RHS was used to measure the hostile content of the Rorschach protocols of all 80 *S*s.

Results

In order to test the relationship between the RHS scores and the ERS scores, it was first necessary to transform the raw scores into standard scores. By comparing the two sets of scores in relation to the variables group judgments of person ("hostile" or "friendly"), self-concept ("hostile" or "friendly"), and kind of interpretation received (“hostile” or “friendly”), 27 possible combinations were arrived at, which are indicated in Table 1.

Since the RHS distribution was quite skewed as compared to a normal distribution, when tested by the chi-square distribution ($p < .01$), a Pearson correlation between the two sets of scores (ERS and RHS) was not computed. Instead, the scores were converted into ranks and correlated by means of Kendall's Tau (6). None of the correlations were significant at the .05 level. This is not unexpected since the use of ranks discards available information. A test for differences in location, however, using Wilcoxon's t for paired replicates (8), resulted in nine significant differences at the .05 level, which for 27 computations was significant beyond the .001 level (9). These significant t's are indicated in Table 2.

The main question arising from the data was, which of the three *groups* of variables, “kind of interpretation” received, “group evaluation” of the *S*, and the “self-concept,” showed the most significant change in the perception of hostility from the Rorschach to the examiner rating situation. It is apparent from examination of Table 2 that the “kind of interpretation” given is involved in the greatest number of significant differences, seven. The group evaluation shows five significant differences, while the self-concept shows four.

The four significant groupings receiving a “friendly” interpretation show that these *S*s perceived significantly less hostility in the examiner than they did on the Rorschach. In other words, as compared to the amount of hostility they projected on the Rorschach, *S*s who were told that they were “friendly” persons, perceived the examiner as being significantly more friendly and competent. The same is true of the three significant differences involving the “friendly” persons. Here, persons who *were friendly*, perceived little hostility in the examiner in comparison to the hostility they perceived on the Rorschach. This fact occurred despite the variation in the kind of interpretation received and in the perception of the “self.” On the other hand, a greater perception of hostility occurred on the ERS than Rorschach for the three “hostile” interpretation groupings and one “hostile,” persons group. Perhaps unexpected is the fact that hostile insightful persons (group evaluation “hostile,” self-concept “hostile”) perceived more hostility on the Rorschach than as a result of ego-threat (examiner situation). The self-concept “friendly” variable showed significant differences when interacting with the “friendly” interpretation variable, with

Table 1

All Possible Combinations (27) of the Variables, Group Interpretation “Hostile,” Group Interpretation “Friendly,” Group Evaluation “Hostile,” Group Evaluation “Friendly,” Self-Concept “Hostile,” and Self-Concept “Friendly”

Number of Subjects	Variables
80	All persons
40	“Hostile” interpretation
40	“Friendly” interpretation
40	Group evaluation “hostile”
40	Group evaluation “friendly”
40	Self-concept “hostile”
40	Self-concept “friendly”
20	Group evaluation “hostile” + “friendly” interpretation
20	Group evaluation “hostile” + “hostile” interpretation
20	Group evaluation “friendly” + “friendly” interpretation
20	Group evaluation “friendly” + “hostile” interpretation
20	Self-concept “hostile” + “friendly” interpretation
20	Self-concept “hostile” + “hostile” interpretation
20	Self-concept “friendly” + “friendly” interpretation
20	Self-concept “friendly” + “hostile” interpretation
20	Group evaluation “hostile” + self-concept “hostile”
20	Group evaluation “hostile” + self-concept “friendly”
20	Group evaluation “friendly” + self-concept “hostile”
20	Group evaluation “friendly” + self-concept “friendly”
10	Group evaluation “hostile” + self-concept “hostile” + “friendly” interpretation
10	Group evaluation “hostile” + self-concept “hostile” + “hostile” interpretation
10	Group evaluation “hostile” + self-concept “friendly” + “friendly” interpretation
10	Group evaluation “hostile” + self-concept “friendly” + “hostile” interpretation
10	Group evaluation “friendly” + self-concept “hostile” + “friendly” interpretation
10	Group evaluation “friendly” + self-concept “hostile” + “hostile” interpretation
10	Group evaluation “friendly” + self-concept “friendly” + “friendly” interpretation
10	Group evaluation “friendly” + self-concept “friendly” + “hostile” interpretation

Table 2

Groupings Showing Significant Differences Between RHS and ERS Scores as Determined by Wilcoxon Paired Replicate *t* Tests

N	Groups	*p*	ERS = perception of hostility higher in ERS than RHS RHS = perception of hostility higher for RHS than ERS
40	"Hostile" interpretation	.05*	ERS
40	"Friendly" interpretation	.05*	RHS
40	Group evaluation "friendly"	.05*	RHS
20	Group evaluation "hostile" + "hostile" interpretation	.02**	ERS
20	Group evaluation "friendly" + "friendly" interpretation	.01***	RHS
20	Self-concept "friendly" + "friendly" interpretation	.01***	RHS
20	Self-concept "friendly" + "hostile" interpretation	.05*	ERS
20	Group evaluation "hostile" + self-concept "hostile"	.05*	RHS
10	Group evaluation "friendly" + self-concept "friendly" + "friendly" interpretation	.01***	RHS

* Significance beyond .05 level.
** Significance beyond .02 level.
*** Significance beyond .01 level.

more hostility being perceived as occurring on the Rorschach than in the examiner. Lastly, those persons whose self concept was "friendly," but who received a "hostile" interpretation, perceived more hostility in the examiner than they did on the Rorschach.

Discussion

The situational quality of projection would seem to be strongly emphasized by the results. Apparently, the objective personality and the self-concept are less important than the kind of situation in which the *S* finds himself. Being told that they were hostile caused the majority of *S*s to project this hostility on to a suitable object (the examiner giving the interpretation) by perceiving him as the "hostile one." A "friendly" report elicited in the main, little perception of the interpreter as hostile.

The situation is not, however, the sole determinant of the operation of projection. Those *S*s who were "friendly" tended to perceive less hostility in the examiner as compared to the Rorschach no matter what kind of interpretation they received nor what their self-concept. The same was not true of the "hostile" *S*s, who did not show any significant difference when considered as an entity. Apparently, the perceptions of "hostile" *S*s were more readily influenced by changes in the environment studied than was true of "friendly" *S*s. Nevertheless, hostile insightful *S*s (group evaluation "hostile," self-concept "hostile") projected considerably more hostility on the Rorschach than in the examiner situation. That this fact is not due to the mere possession of "hostility" may be gleaned from the fact that the group evaluation "hostile," self-concept "friendly" group did not show any significant difference in their perception of hostility in the two situations. It seems safer to conclude, therefore, that the projection of hostility on the Rorschach in this study was a function of the possession of hostility in conjunction with insight into this fact. Moreover, the fact that "friendly" *S*s (without considering their self-concepts) tended to manifest more hostile perceptions on the Rorschach than they did with regard to the interpreter seems at variance with some of the current thought about the meaning of responses elicited by projective techniques, namely, that elicited responses are representative of a person's phenomenological world. Apparently, clinical psychologists might well take note of the fact that elicited hostile responses on the Rorschach may at times reflect only the person's freedom from fear in expressing that which is indicative of his "life style." Conversely, the absence of hostile responses in a protocol may for some persons be indicative not of nonhostile feelings, but rather of a reluctance or fear of expressing strong hostile feelings through perception on the Rorschach. Indeed, from a psychotherapeutic point of view, those persons who are hostile but deny this fact (hostile—noninsightful) might be considered to be more deeply disturbed than those who are hostile, possess insight into this fact, and are willing to acknowledge their possession of hostility openly. Seemingly therefore, a revision of the rationale underlying the Rorschach is needed.

The concept of one's "self" as friendly is of importance in interaction with other variables. It is the only variable that showed a shift in the perception of the examiner from "friendly" to "hostile" when the kind of interpretation given was changed from "friendly" to "hostile." This serves to confirm the stand that these persons were operating from a phenomenological frame of reference, in that when the interpretation given was in accordance with the self-concept ("friendly" interpretation, "friendly" self concept) there was little perception of hostility on the part of the examiner. When the interpretation differed from the self concept ("hostile" interpretation, "friendly" self concept), the perceived threat to the "self" resulted in the projection of the *S*'s hostility on to the interpreter.

Summary

The results seem to indicate that *S*s projecting hostility in one situation (Rorschach) show significant differences from the tendency to perceive hostility in another situation (reaction through rating of the examiner's interpretation of their protocol). These differences have been shown to be related to such variables as kind of interpretation of their protocol given to the *S*s by the interpreter, group evaluation of the *S* by his fraternity brothers, and the *S*'s self concept. Although each of these variables was related to the perception of hostility, the strongest determinant appeared to be the emotional climate created by the interpretation given to the *S*s regarding their Rorschach performance. Those *S*s called "hostile" reacted to this threat to the "self" by rating the interpreter as "hostile." Those *S*s called "friendly" rated the interpreter as competent and "friendly."

An additional finding was that "friendly" *S*s tended to perceive more hostility on the Rorschach than they did in the examiner rating situation. This finding was not true of "hostile" *S*s whose perceptions were influenced by whether they received a "friendly" or "hostile" interpretation of their Rorschach protocols. Lastly, the finding that "friendly" *S*s, and "hostile-insightful" *S*s showed a significantly greater production of hostile responses on the Rorschach than in the examiner rating situation seems to warrant further investigation as to the meaning of responses obtained on the Rorschach.

Received April 4, 1957.

References

1. Cattell, R. B. Principles of design in "projective" or misperception tests of personality. In H. H. Anderson & Gladys L. Anderson (Eds.), *An introduction to projective techniques*. New York: Prentice-Hall, 1951. Pp. 55–98.
2. Elizur, A. Content analysis of the Rorschach with regard to anxiety and hostility. *J. proj. Tech.*, 1949, **13**, 247–284.
3. Gluck, M. R. The relationship between hostility in the TAT and behavioral hostility. *J. proj. Tech.*, 1955, **19**, 21–26.
4. Gluck, M. R. Rorschach content and hostile behavior. *J. consult. Psychol.*, 1955, **19**, 475–478.
5. Murstein, B. I. The projection of hostility on the Rorschach, and as a result of ego-threat. *J. proj. Tech.*, 1956, **20**, 418–428.
6. Smith, K. Distribution-free statistical methods and the concepts of power efficiency. In L. Festinger & D. Katz (Eds.), *Research methods in the behavioral sciences*. New York: Dryden Press, 1953. Pp. 536–577.
7. University of Texas Interview Rating Scale (Form A) Testing and Guidance Bureau (mimeographed). Austin: Univer. of Texas, 1955.
8. Wilcoxon, F. *Some rapid approximate statistical procedures*. New York: American Cyanamid, 1949.
9. Wilkinson, B. A statistical consideration in psychological research. *Psychol. Bull.*, 1951, **48**, 156–158.

of choice and social desirability obtained in this study is comparable to the correlation of .87 reported by Edwards (2) when his items were administered in the "Yes-No" type of inventory. While the 30 item pairs upon which the present correlation is based is relatively small, the correlation appears to be more likely the result of the method used to judge social desirability rather than the result of a biased or restricted sample. It would seem that, from responses to single items, one cannot generalize with impunity about the influence of social desirability on forced-choice pairings. Our findings suggest that paired items can acquire contextual meaning which alters the values assigned to the same items when responded to separately. They are consistent with other investigations, e.g., Howes and Osgood (5), which demonstrate the influence of context in changing the meaning of individual words. It appears likely that Edwards, in using the method of successive intervals to scale his items, achieved only a first approximation of equal pairings and that additional judgments of the item pairs themselves are necessary along with revisions in pairing before the variable of social desirability can be eliminated from the EPPS.

Summary

The present study investigated the extent to which social desirability influences responses to a short form of the *Edwards Personal Preference Schedule*. The results indicate that the factor of social desirability is still an important influence since, in a statistically significant number of item pairs, judges had a definite preference for one member of the pair as being more desirable. Also, the preferences of judges in terms of social desirability correlated highly with actual choices of another group of *S*s who took the scale as a standard personality test.

Received October 21, 1957.
Early Publication.

References

1. Edwards, A. L. The scaling of stimuli by the method of successive intervals. *J. appl. Psychol.*, 1952, **36**, 118–122.
2. Edwards, A. L. The relationship between judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.*, 1953, **37**, 90–93.
3. Edwards, A. L. *Edwards Personal Preference Schedule*. Manual. New York: Psychological Corp., 1954.
4. Edwards, A. L., & Thurstone, L. L. An internal consistency check for scale values determined by the method of successive intervals. *Psychometrika*, 1953, **17**, 169–180.
5. Howes, D. H., & Osgood, C. E. On the combination of associative probabilities in linguistic contexts. *Amer. J. Psychol.*, 1954, **67**, 241–258.
6. Klett, C. J. The stability of the social desirability scale values in the Edwards Personal Preference Schedule. *J. consult. Psychol.*, 1957, **21**, 183–185.
7. Sakoda, J. M., Cohen, B. H., & Beall, G. Test of significance for a series of statistical tests. *Psychol. Bull.*, 1954, **51**, 172–175.

Journal of Consulting Psychology
Vol. 22, No. 1, 1958

A Factorial Isolation of Psychiatric Outpatient Syndromes[1]

Mary Helen Tatom
Legal Psychiatric Services, D. C. Department of Public Health

This study represents an attempt to validate certain psychiatric diagnostic entities, using the technique of obverse factor analysis to establish modal patterns of symptom groupings. If clinical classifications of patients are valid, they should divide pathological traits into syndromes. It was the problem of this paper to test the existence of four distinct disease reactions corresponding to psychiatric diagnoses in an outpatient population, and to establish by means of the technique of obverse factor analysis patterns of characteristics common to the members of each group and not possessed by members of other groups.

Procedure

Subjects

The *S*s were 20 male veterans who were receiving psychotherapeutic treatment in the Veterans Administration Mental Hygiene Clinic of the Washington, D. C., regional office. Their ages ranged from 25 to 44 years, with a median age of 32 years. Five *S*s were selected from each of the following four diagnostic categories: Hysteric, Anxiety State, Schizophrenic, and Obsessive-Compulsive. At least two medical diagnoses were used in the selection of the psychoneurotic *S*s. The therapist's diagnosis was confirmed by a concurring diagnosis from another source.

A somewhat different procedure was followed in the selection of schizophrenic *S*s. Since these *S*s had not been hospitalized, and prodromal symptoms had in some instances been previously described as an "anxiety state," no confirming medical diagnosis was secured on these patients. They were selected on the basis of the current diagnosis of their respective treating physicians plus case history.

Measures

The measures consisted of 67 rating scales which were essentially the same as those analyzed by Lorr and Rubinstein (9) in an R-type factor analysis using a population of psychiatric outpatients. The following are examples of the kind of traits which the scales were designed to measure: Emotional Reactivity, Concern with Conformity, Hostile Passive Obstructionism, Compulsions, and Gas-

Table 1

Oblique Factor Matrix $V = F \Lambda$

Variable	Factor A	Factor B	Factor C	Factor D
1. H-1	.07	.29	.31	−.02
2. H-2	−.26	−.11	−.02	−.49
3. H-3	.02	.01	.69	−.03
4. H-4	.13	.00	−.04	−.49
5. H-5	−.23	.44	.13	.02
6. A-1	−.20	.49	.11	−.01
7. A-2	−.08	.44	.58	.26
8. A-3	−.29	.33	−.17	−.10
9. A-4	.00	.57	.01	.04
10. A-5	−.25	.19	.40	.29
11. S-1	.20	−.17	−.03	−.36
12. S-2	.38	.13	.38	.12
13. S-3	.32	.14	.27	.07
14. S-4	.05	.29	−.26	−.03
15. S-5	.16	.19	−.22	.07
16. O-1	.39	.16	−.44	−.17
17. O-2	.37	−.07	−.04	−.04
18. O-3	.60	.07	−.07	−.03
19. O-4	.58	−.01	−.36	−.17
20. O-5	.66	−.03	.00	−.05

---

[1] This article represents a treatment of one of two problems included in a doctoral dissertation, Catholic University of America. It was carried out using the facilities of the Veterans Administration Mental Hygiene Clinic, Washington, D. C.

Table 2

Trait Composition of Primary Factors

Scale No.	Scale Description	Primary Factors A	B	C	D
1	Underreacts emotionally—overreacts	+	−	−	
2	Depressed—elated	+		−	
3	Even in mood—mood swings	+		−	
4	Inhibited—uninhibited	+	−	+	+
5	Tense—phlegmatic	+		+	−
6	Concerned with self—unconcerned				+
7	Lacking self-assurance—self-assured	+		−	
8	Selfconscious—unselfconscious	+		−	+
9	Submissive—dominant	+		−	
10	Rigid—flexible	+		−	
11	Outgoing—seclusive	−		+	
12	Poorly adjusted socially—gains acceptance	+		−	
13	Rarely sees a task through—perseveres	−	+	−	
14	Effeminate—masculine		+		−
15	Unsatisfied need for affection—no unsatisfied need		+		
16	Egocentered—nonegocentered	+	+	−	
17	Seeks support from others—does not seek support				+
18	Lacks initiative—acts independently			−	
19	Avoids responsibility—accepts responsibility	+	+	−	−
20	Feels rejected—feels accepted	+	+	+	
21	Perceives world as hostile—perceives world as friendly	+			
22	High level of achievement—low level of achievement	−			+
23	Considers the future—does not consider the future		−	+	
24	Strong superego—weak superego		−		+
25	Impulse control—lack of impulse control		−	+	
26	Orderly—disorderly	+	−		+
27	Perfectionistic—unconcerned with tasks		−	+	+
28	Concerned with conformity—nonconformist		−	+	+
29	Hostile to authority—accepts authority	+			−
30	Blames self—does not blame self	−	−	+	
31	Projects blame—does not project blame	+	+		−
32	Denies the existence of difficulties—exaggerates difficulties	−		−	+
33	Has insight—lacks insight		−		+
34	Overt hostility—no overt hostility	−			
35	Hostile withdrawal—no hostile withdrawal	+	−		
36	Hostile passive obstructionism—no hostile passive obstructionism		+	−	+
37	Interested in activities around him—lacks interest	−		+	
38	Frustration tolerance—lacks frustration tolerance		−	+	
39	Needs approval—does not need approval				+
40	Suspicious—trustful	+			
41	Concerned over hostility—unconcerned				+
42	Overconcerned with symptoms—unconcerned with symptoms		+	−	−
43	Autism—no autism	+	+		
44	Anxious—unworried	+	+	+	
45	Hostile—not hostile	+	+	−	
46	Irritable—not irritable		+	−	+
47	Unrealistic ambition—no unrealistic ambition	+	−	−	−
48	Sex conflict—no sex conflict				+
49	Uses complaints—no use of complaints	+		−	
50	Unreality feelings—no unreality feelings			−	+
51	Phobias—no phobias	+	−		+
52	Suicidal thoughts—no suicidal thoughts	−	+	−	−
53	Reality distortion—no reality distortion	+	−	−	
54	Ideas of ıeference—no ideas of reference	+	−	−	

Table 2—*Continued*

Scale No.	Scale Description	Primary Factors A	B	C	D
55	Obsessive thoughts—no obsessive thoughts			−	+
56	Obsessive hostile impulses—no obsessive hostile impulses				−
57	Compulsions—no compulsions	−	−	−	
58	Alcoholic addiction—no alcoholic addiction	−	+	+	
59	Marked behavioral reaction to alcohol—no behavioral reaction to alcohol				−
60	Conversion symptoms—no conversion symptoms		+		
61	Anergic complaints—no anergic complaints		+	+	
62	Body preoccupation—no body preoccupation				−
63	Headaches—no headaches	+	+		−
64	Sleep difficulties—no sleep difficulties	+			
65	Intestinal symptoms—no intestinal symptoms	+	−		−
66	Gastric symptoms—no gastric symptoms	−	−	+	−
67	Cardiovascular symptoms—no cardiovascular symptoms	−	−		+

tric Symptoms. The form of the rating scales is illustrated below:

Trait 28. Concern with Conformity:

1. Shows excessive and chronic concern for conformity with standards of the group. Would hesitate long before committing a violation.
2. Shows a fairly strong concern over adherence to social conventions. Feels uncomfortable if he is not conforming.
3. Shows a slightly more than average concern over adherence to social conventions.
4. Shows somewhat less than average concern over adherence to social conventions.
5. Shows little concern over disregard of social conventions.
6. Shows no concern over disregard of social conventions.

X. Unrateable. (Specify)

Ratings were secured from each patient's individual therapist. Since the patients had been in therapy for varying lengths of time (one month to two years), an attempt was made to control the effect of changes that might have taken place during the therapeutic process by instructing the therapist to refer to the notes of his earliest sessions in rating those patients who, he felt, had changed under treatment. The data analyzed included ratings by psychiatrists, clinical psychologists, and social workers. To insure comparability of ratings, a rating guide was distributed to each person responsible for evaluating a patient. Further clarification was obtained by devoting a two-hour staff conference in the participating clinic to a discussion of scale items.

Table 3

Oblique Reference Vectors in Relation to Unrotated Factors

Unrotated Factors	Rotation Matrix Λ Reference Vectors A	B	C	D
I	.48	−.21	−.36	.02
II	.18	.28	−.43	−.41
III	.76	.19	.57	−.18
IV	−.40	.92	.60	.89

Table 4

Relations of Oblique Reference Vectors to One Another
Λ′Λ-Direction Cosines

Vector	A	B	C	D
A				
B	−.28			
C	−.05	.61		
D	−.55	.67	.60	

Table 5

Relations of Primary Factors to One Another
D $(\Lambda'\Lambda)^{-1}$ D

Factor	T_a	T_b	T_c	T_d
T_a				
T_b	.00			
T_c	−.40	−.32		
T_d	.60	−.38	−.48	

Table 6

Unrotated Factor Matrix F_s[a]

Primary Factor	Second-order Factor	
	I_s	II_s
A	.707	.236
B	−.192	.698
C	−.461	−.498
D	.911	−.231

[a] Subscript indicates second-order factors.

Design

The technique of obverse factor analysis was used to determine whether the four diagnostic entities could be reproduced statistically and also to determine any inherent relations among these so-called entities. The resulting factors might be considered as representing types of pathological adjustment, which could further be described in terms of the individual traits they comprise.

Tetrachoric correlations were obtained among the 20 *S*s, 5 from each diagnostic category. Traits were dichotomized at the interval nearest the median of the 20 patients. Using Thurstone's (12) centroid method of factor analysis, loadings on four factors were extracted from the 20 × 20 correlation matrix. Factor extraction was stopped when the distribution of residual *r*'s around zero approximated a chance distribution. Rotation to simple structure was carried out according to the method described by Thurstone (12, pp. 194–216). Simplicity of structure was given priority over orthogonality. The correlations between primary vectors were determined and the resulting matrix factored, yielding two second-order factors. These were rotated so as to retain an orthogonal structure. The factors arrived at statistically were examined as to the personality pattern they delineate and compared with present psychiatric concepts of disease entities.

Table 7

Rotated Factor Matrix $V_s = F_s\Lambda_s$

Primary Factor	Second-order Factor	
	X	Y
A	.530	.001
B	−.026	.514
C	−.422	−.233
D	.667	−.359

Table 8

Rotation Matrix Λ_s

Factor	X	Y
I_s	.675	.223
II_s	−.223	.675

Results

The Rotated (oblique) Factor Matrix is given in Table 1.[2] The following procedure was used in identifying primary factors: three individuals were selected as representatives of each factor, taking into consideration both saturation with that factor and relative purity, freedom from loadings on other factors. Factor A was represented by individuals O-3, O-4, and O-5; Factor B by H-5, A-1, and A-4; Factor C by H-3, A-2, and A-5; and Factor D by H-2, H-4, and S-1. A trait was considered as contributing to a factor when all of the three representatives of the factor fell on the same side of the median in the trait as rated. The composition of the factors in terms of traits derived by this procedure is presented in Table 2. A pattern of distin-

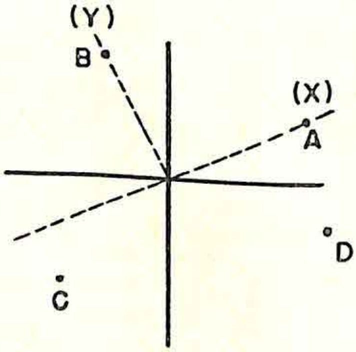


Fig. 1. Rotation of second-order factors.

---

[2] Correlation and centroid factor matrices may be obtained without charge from Mary H. Tatom, 3801 Connecticut Ave., Washington 8, D. C., or from the American Documentation Institute. Order Document No. 5434, remitting $1.25 for microfilm or $1.25 for photocopies.

Table 9

Factor X

Scale No.	Scale Description: Positive	Negative
2	Depressed	vs. Elated
3	Even mood	vs. Mood swings
7	Conspicuously lacking in self-assurance	vs. Self-assured
9	Submissive	vs. Dominant
10	Rigid	vs. Flexible
11	Seclusive	vs. Outgoing
12	Poorly adjusted socially	vs. Gains acceptance socially
37	Little interest in activities around him	vs. Interested in activities around him
49	Uses complaints	vs. Does not use complaints

guishing traits for each primary factor was derived from Table 2 by assigning to each primary factor those traits which differentiated it from all others, i.e., those traits which were not represented at all in any other factor, or which were not represented similarly (as to presence or absence) in another factor. Further discussion of the trait composition of primary and second-order factors will be in terms of these distinguishing (pathognomonic) traits.

The cosine matrix (Table 5) of the primary factors was factored to obtain second-order parameters. The two second-order factors were plotted as a step toward their orthogonal rotation. Inspection of this plot shown in Fig. 1 indicates certain interesting relationships. The rotation indicated by dotted lines in Fig. 1 and as completed in Table 7, permitted a tentative identification of Factor X by examining those traits which differentiated between Factors A and C and were in no way involved in Factor B. Traits which differentiated between Factor C and Factor D were included if they were not involved in Factor B. Second-order Factor Y was tentatively identified by isolating those traits which were peculiar to B (for which there was no indication in other factors) or which differentiated between B and D and did not differentiate between A and C. Of this group of traits, those which showed some relationship between D and either A or C were omitted, since D had some loading on X. Traits of second-order Factors X and Y are given in Tables 9 and 10, respectively.

Table 10

Factor Y

Scale No.	Scale Description: Positive	Negative
4	Uninhibited	vs. Inhibited
13	Rarely sees a task through	vs. Perseveres
14	Effeminate	vs. Masculine
15	Unsatisfied need for affection	vs. No unsatisfied needs to be loved
24	Lack of superego	vs. Severe superego
26	Disorderly	vs. Orderly
27	Unconcerned with tasks	vs. Perfectionistic
28	Does not conform	vs. Concerned with conformity
33	Lack of insight	vs. Insight
42	Overconcerned with symptoms	vs. Not concerned with symptoms
51	No phobias	vs. Phobias
52	Suicidal impulses	vs. No suicidal impulses

Discussion

The factors isolated statistically divided patients into groups which cut across clinical diagnostic classifications. The clinically diagnosed *Obsessive-Compulsives* clearly defined Factor A. Factor B showed heaviest loadings in two patients who had been clinically diagnosed as *Anxiety State* and one patient who had been diagnosed *Hysteric*. The three individuals with highest loadings in Factor C had been diagnosed as *Anxiety State* (two patients) and *Hysteric* (one patient). The negative pole of Factor D was defined by two persons who had been diagnosed as *Hysteric* and one who had been diagnosed as *Schizophrenic*. In terms of traits, the syndromes derived by statistical analysis have been tentatively related to clinical concepts as follows:

Primary factors
- Factor A—Outpatient paranoid schizophrenic
- Factor B—Conversion hysteric
- Factor C—Intrapunitive, socially mature personality vs. narcissistic personality
- Factor D—Passive dependent personality vs. anti-social personality

Second-order factors
- Factor X—Schizothymia vs. cyclothymia
- Factor Y—Uncontrolled emotionality vs. overcontrolled emotionality

Factor A

It will be observed that traits specific to A, not involved in Factor X, are traits implying reality distortion, flattened affect, and paranoid ideation:

Trait No.	
1	Underreacts emotionally.
21	Perceives world as hostile.
34	No overt hostility.
40	Suspicious.
47	Unrealistic ambition.
53	Reality distortion.
54	Ideas of reference.
58	No alcoholic addiction.
64	Sleep difficulties.
65	Intestinal symptoms.

This symptom pattern appears to be classically schizophrenic. Clinically diagnosed obsessive-compulsive neurotics formed a homogeneous group on the basis of the pattern of traits represented by Factor A, while clinically diagnosed schizophrenics showed relatively low loadings on all factors. The finding that clinically diagnosed schizophrenics had little common factor variance in this analysis suggests that certain essentially schizophrenic characteristics were not covered by the rating scales used. It is possible that paranoid schizophrenia developing in an obsessive-compulsive personality structure presents a consistent pattern, while other types of incipient schizophrenia show such varied overt symptomatology that early schizophrenia as such cannot be isolated as a clear-cut entity. Clinicians are familiar with the wide variety of behavior on which the diagnosis of incipient schizophrenia may be based: panic states, bizarre acting out behavior, a definite change in personality, and general intellectual or behavioral disorganization. The patients who define Factor A resemble ambulatory schizophrenics whose adjustment is stabilized at the level of minimal social functioning, who retain some controls, and who are not at present threatened with an acute psychotic episode. Such patients might receive the diagnosis of obsessive-compulsive neurotic on the basis of characterological features such as rigidity, inhibition, and absence of overtly hostile behavior, though neurotic symptoms were not present.

Factor A is somewhat similar to Lorr and Rubinstein's (9) Factor B and (10) Factor G, and to the negative poles of Cattell's (4) Factors A, H, and L.

Factor B

Factor B is composed of a group of characteristics descriptive of unrestrained emotional behavior. The syndrome isolated here corresponds to some extent to the General Emotionality factors of Burt (2), Eysenck (6), Rao and Slater (11), and Cattell (4). In the present study, poor emotional control, an unsatisfied need for love and care, and lack of insight are associated with conversion symptoms. In terms of psychiatric formulations, the group of traits which make up Factor B are those usually associated with conversion hysteria. The patients who defined this factor had been diagnosed Anxiety State (two patients) and Hysteric (one patient).

Factor C

Factor C is a bipolar factor, described at its positive pole by a group of patients who are socially adjusted, independent, and nonegocentered. On the pathological side, they tend to assume blame and present gastric symptoms. Perhaps in a clinic population, independence and cooperativeness are achieved at the expense of inner strain. The negative pole of Factor C is characterized by a group of traits which involve a hostile, passive, and egocentered adjustment which corresponds to the clinical concept of narcissistic character disorder. The diagnosis *Inadequate Personality* is sometimes applied to this type of patient. Though the negative pole of the factor is not defined by a group of patients, two of the clinically diagnosed obsessive-compulsives show significant negative loadings on Factor C (using .35 as a criterion of significance). The patients who defined Factor C had been clinically diagnosed Anxiety State (two patients) and Hysteric (one patient).

Traits which are specific to Factor C which do not appear in Factor X are as follows:

16 Nonegocentered vs. egocentered.
18 Acts independently vs. lacks initiative.
30 Blames self vs. does not blame self.
36 No hostile passive obstructionism vs. hostile passive obstructionism.
45 Not hostile vs. hostile.
46 Not irritable vs. irritable.
66 Gastric symptoms vs. no gastric symptoms.

The dimension involved in Factor C appears to be basically one of unselfishness vs. selfishness. Factor C resembles somewhat Lorr and Rubinstein's (9) Factor E, which they termed dependent immaturity vs. independent maturity. It is also similar to Cattell's (4) Source Trait A. Factor C corresponds in general to the surface trait which Cattell (4) describes as infantile, demanding self-centeredness vs. emotional maturity, frustration tolerance.

Factor D

Factor D cannot be definitely described as a bipolar factor. However, since it is defined by a group of patients at its negative pole, positive loadings approach significance (.29), and the factor loadings in both directions are relatively low (− .49 to .29), characteristic of both poles are presented.

Traits specific to D, not involved in second order factors are as follows:

5 Phlegmatic—tense.
6 Self preoccupied—attentive to external matters.
17 Seeks support from others—self-sufficient.
31 Does not project blame—projects blame.
32 Denies frustration—exaggerates frustration.
39 Needs approval—does not need approval.
41 Concerned over hostility—unconcerned over hostility.
48 Sex conflict—no sex conflict.
56 No obsessive hostile impulses—obsessive hostile impulses.
59 No behavioral reaction to alcohol—marked behavioral reaction to alcohol.
62 No body preoccupation—body preoccupation.
63 No headaches—headaches.
67 Cardiovascular symptoms—no cardiovascular symptoms.

This factor presents a syndrome characterized at its positive pole by submissiveness, inner conflict, insecurity, guilt, and dependency and at its negative pole by conflict with authority, aggression, absence of guilt, and self-sufficiency. Cardiovascular symptoms are associated with the former and headaches with the latter. Introversion, guilt, conflict, and inhibition describe a personality structure associated with the diagnosis of anxiety neurosis, and cardiovascular symptoms may perhaps be an autonomic expression of anxiety. The factor under consideration has been related to the clinical diagnosis *Passive Dependency Reaction* because of the absence of overt anxiety symptoms; however, inner conflict and inadequacy feelings appear to be central to this syndrome, one of extreme inner discomfort. The basic dimension of Factor D is one of submission to outer authority and to superego dictates vs. rebellion against outer authority and absence of guilt. The cluster of traits which characterizes the negative pole of Factor D suggests a psychopathic deviate syndrome. The dimension represented by Factor D appears to be very similar to Eysenck's (6) factor which divided his population into a social misfit vs. a psychological conflict group. It also corresponds to some degree to Rao and Slater's (11) factor, obsessive-compulsive vs. psychopath. Lorr and Rubinstein's Factor D, which they described as a dimension of inadequacy feelings, is very similar to Factor D. In both factors, headaches are associated with absence of conscious inferiority feelings. Cattell's (3) Source Trait E, ascendance vs. submission appears also to be comparable to Factor D.

Factor X

This second-order factor has been termed a schizothymia vs. cyclothymia dimension in terms of the traits of Factors A and C which define its positive and negative poles respectively. In factorial studies, a schizothyme-cyclothyme factor has frequently appeared, and each pole has appeared as a separate factor at times. Cattell (3) concludes that this dimension represents a basic temperamental difference in people, corresponding to Kretschmer's dichotomy based on body type.

Factor Y

Factor Y represents a dimension defined at its positive pole by primary Factor B. Factor Y appears to be similar to the negative pole of Cattell's (4) Factor C and to the factor General Emotionality found by Eysenck (5), by Rao and Slater (11), and by numerous other analysts.

Perhaps the fact that clinical groupings of patients were lost in the statistical analysis can be explained to some extent by the observation that obverse factor analysis may classify individuals into types on the basis of underlying generalities rather than on readily observable surface characteristics. In an obverse analysis of four types of geometrical solids (cones, rectangular prisms, rectangular pyramids, and cylinders), Lorr, Jenkins, and Medland (8) found that rotation yielded two bipolar composite factors: (*a*) tapering and curvilinear vs. nontapering and rectilinear; and (*b*) tapering and rectilinear vs. nontapering and curvilinear. These represent underlying abstractions rather than physically distinct principles of classification. The distinguishing gestalts of the four geometrical solids did not emerge as factors, nor could they be derived from the factors. In other words, the clusters of geometric solids could be named only from sensory information. Bom Mo Chung (1) reported results comparable to those of the present investigation in a study in which he attempted to arrive at genetic species by an obverse factor analysis of the correlations between species based on 14 genetic traits. He found only rough coincidence of factorial types with conventional genetic species, but felt that this did not invalidate the technique, since factorial types might furnish a valid basis of classification independent of other classifications of the data. Lorr and Fields (7) in a factorial study of body types found Sheldon's three components to represent patterns which could be accounted for by two factors best described as muscle or fat, leg length or torso length.

Primary factors in the present analysis are highly intercorrelated; perhaps the addition of other traits might define them more clearly. It is felt that second-order factors were clearly established in the present analysis, but that primary factors require cross-validation. Primary factors were so heavily saturated with second-order factors that clusters of patients may have been formed largely in terms of only two dimensions, obscuring some clinically observable differences between patients and accounting to some extent for discrepancies between psychiatric and clinical classifications.

The results of the present study cannot be said to validate or invalidate either factorial classification of patients or their classification by conventional psychiatric nosology. It does, perhaps, explain to some extent the frustration involved in attempting to validate psychological measures against clinical diagnoses.

Summary

The present investigation represents an attempt to test by the technique of obverse factor analysis the existence of four distinct disease reactions corresponding to psychiatric diagnoses in an outpatient population and to establish patterns of traits characteristic of each syndrome.

Factorially isolated syndromes cut across clinical diagnostic lines, separating patients into groups which were different from the categories into which they had been placed by clinical diagnosis. The lack of agreement between factorial and psychiatric groupings of individuals was not thought to invalidate either system of classification.

Traits were grouped by this study into aggregations which formed recognizable clinical entities and corresponded in general to factors isolated in other factor analyses. Primary and second-order factors were tentatively identified as follows:

Primary factors
- Factor A—Outpatient paranoid schizophrenic.
- Factor B—Conversion hysteric.
- Factor C—Intrapunitive, socially mature personality vs. narcissistic personality.
- Factor D—Passive dependent personality vs. anti-social personality.

Second-order factors
- Factor X—Schizothymia vs. cyclothymia.
- Factor Y—Uncontrolled emotionality vs. overcontrolled emotionality.

Received March 25, 1957.

References

1. Bom Mo Chung. Differentiation of group patterns by transposed factor analysis. *Ewah Woman's Univer. Press Stud. Psychol.*, Seoul, Korea, 1954, No. I.
2. Burt, C. The factorial study of temperamental traits. *Brit. J. Psychol.*, 1948, **1**, 178–203.
3. Cattell, R. B. *The description and measurement of personality*. Yonkers, N. Y.: World Book, 1946.
4. Cattell, R. B. *Personality*. New York: McGraw-Hill, 1950.
5. Eysenck, H. J. *Dimensions of personality*. London: Kegan Paul, French, & Trubner, 1947.
6. Eysenck, H. J. Types of personality: A factorial study of seven hundred neurotics. *J. ment. Sci.*, 1944, **90**, 851–861.
7. Lorr, M., & Fields, U. A factorial study of body types. *J. clin. Psychol.*, 1954, **10**, 182–185.
8. Lorr, M., Jenkins, R. L., & Medland, F. F. Direct versus obverse factor analysis: A comparison of results. *Educ. psychol. Measmt*, 1955, **15**, 441–449.
9. Lorr, M., & Rubinstein, E. A. Factors descriptive of psychiatric outpatients. *J. abnorm. soc. Psychol.*, 1955, **51**, 514–522.
10. Lorr, M., & Rubinstein, E. A. Personality patterns of neurotic adults in psychotherapy. *J. consult. Psychol.*, 1956, **20**, 257–263.
11. Rao, C. R., & Slater, P. Multivariate analysis applied to differences between neurotic groups. *Brit. J. Psychol. Statist. Sec.*, 1949, **2**, 17–29.
12. Thurstone, L. L. *Multiple-factor analysis*. Chicago: Univer. of Chicago Press, 1947.

PSYCHOLOGICAL TEST REVIEWS

New Test

Lorge, Irving, & Thorndike, Robert L. *Lorge-Thorndike Intelligence Tests.* Grades kgn.-12. 5 levels, 2 forms at each level. Primary Battery (levels 1–2), (30) min., test booklet ($3.00 per 35), with key, record sheet, and manual, pp. 20. Verbal Battery (levels 3–5), 34 (45) min., or Nonverbal Battery (levels 3–5), 27 (40) min. Consumable edition: test booklet ($2.40 per 35), with key, record sheet, and manual, pp. 32. Re-usable edition: test booklet ($2.40 per 35), with manual; IBM answer sheet ($1.20 per 35), with record sheet; IBM or hand-scoring key (21¢ ea.). Technical manual, pp. 16 (12¢). Boston: Houghton Mifflin, 1957.

The Lorge-Thorndike Intelligence Tests comprise eight different tests which span the range from kindergarten through high school. Each is available in two equivalent forms, and the three higher levels are furnished in two formats for use with or without separate answer sheets. For the Primary Battery, on two levels covering kindergarten and the first three grades, administration is oral and responses are made by marking pictures. Three types of items—oral vocabulary, pictorial classification, and pictorial pairing—are used. Levels 3 to 5, covering the range from Grade 4 to Grade 12, are provided with verbal and nonverbal batteries in separate booklets. The subtests of the Verbal Battery are word knowledge, sentence completion, verbal classification, verbal analogies, and arithmetic reasoning. The Nonverbal Battery has figure analogies, figure classification, and number series.

The Technical Manual shows that the development and standardization of the tests involved the examination of more than 136,000 children in 44 communities in 22 states. Norms were based on a stratified sampling of communities rated in five groups by socioeconomic criteria. For all levels and tests, four types of norms may be read directly from raw scores: deviation IQs ($M = 100$, $SD = 16$), grad percentiles, grade equivalents, and age equivalents Other technical data are fully presented and impressive. Alternate form reliabilities in single-grade samples range from .76 to .89 with, not unexpectedly, the higher reliabilities appearing in the verbal tests at the higher levels. Corrected split-half reliabilities for single-grade groups run from .88 to .94. Verbal and nonverbal batteries correlate about .65. A factor analysis of the subtests shows that verbal and nonverbal factors account for most of the variance. Congruent validities with other appropriate tests are about .65, and the one reported study of predictive validity shows a correlation of scores and subsequent ninth-grade achievement to be .67.

These tests have many merits in their well-planned and attractive formats and their numerous conveniences for use. But their greatest strength is the homogeneous content and the common standardization over so wide a range of developmental levels. They will give excellent service for many applications to educational administration, classification, and guidance.—*L. F. S.*

Journal of Consulting Psychology
Vol. 22, No. 2, 1958

Relationship Between Certain Personality Variables and Continuation in Psychotherapy[1]

Earl S. Taulbee
VA Mental Hygiene Clinic, Omaha, Nebraska

It is obvious from the wide range of variables (pertaining to patient, therapist, technique, environment, and the interaction of these) which have been found to be associated with continuation and improvement in psychotherapy, that the process and outcome of treatment are extremely complex. In general, studies reported to date offer support for the relationship between such factors as socioeconomic status, education, intelligence, and number of interviews to continuation in psychotherapy and to its outcome. Very little research has been done on the role of specific personality variables in the process of treatment. Because of the limited facilities available for treatment, accurate prognosis is a very important problem.

Cartwright (2) studied the outcome of 78 clients seen in client-centered therapy. He interpreted the relation found between success ratings and length of therapy as suggesting that certain individual differences between clients give rise to different kinds of therapeutic process. Two kinds of process were identified as: "short" (1–12 interviews) and "long" (13–77 interviews). Very similar results were reported by Taylor (9) using cases seen in a psychoanalytically oriented VA mental hygiene clinic. These two studies offer some evidence that continuation and rated improvement in therapy are more a function of personality differences in clients than the type of therapy employed. Gallagher (4) found that the more anxious the individual the more likely he would be to remain in therapy and the more defensive the individual the greater the likelihood of his prematurely leaving therapy. Investigating the differences between a group of patients rated improved and an unimproved group, Rosenberg (6) found that the improved showed higher IQ, greater productivity, greater emotional depth and responsiveness, more sensitivity and tact, and higher energy level and drive for achievement. The unimproved showed more stereotypy and more preoccupation with physical complaints. Mowrer suggested that ". . . depression, inferiority feeling, anxiety, and other painful affects which the neurotic experiences are *signs of strength* within the personality . . . an individual with internal disharmony who is experiencing these distressing emotions is not nearly so badly off as he would be if, under the same circumstances, he were bland, unconcerned, unmotivated" (5, p. 91).

Purpose of the Present Study

It is the author's opinion that certain identifiable personality variables are associated with the premature termination of, or continuation in, individual psychotherapy. Furthermore, regardless of the way in which these interact with other important variables in the therapeutic relationship, they will be reflected in certain psychological tests. The following specific hypotheses were tested:

Hypothesis 1. Certain symptoms and affects, as revealed by specific MMPI scales and Rorschach variables, are positively related to continuation in psychotherapy:

(*a*) The "continuers" show a greater elevation of the "symptom" scales of the MMPI (*Hs, D, Pa, Pt,* and *Sc*) than the "attriters."

(*b*) The continuers give a larger number of color (*C*), shading (*Y*), vista (*V*), and anatomy (*An*) responses on the Rorschach

---

[1] From the Veterans Administration Mental Hygiene Clinic, Omaha, Nebraska.

Table 1

Comparison of Continuer and Attriter Groups on MMPI Scales

MMPI Scale	Continuers Mean	Continuers *SD*	Attriters Mean	Attriters *SD*	*t*	*p**
?	8.27	11.13	8.08	10.70	.08	N.S.
L	3.60	1.99	4.50	2.59	1.79	N.S.
F	6.02	3.91	5.13	3.16	1.13	N.S.
K	12.38	5.16	12.98	4.89	.54	N.S.
Hs	80.36	15.46	77.38	11.57	.98	N.S.
D	78.29	15.47	70.43	12.40	2.53	.025
Hy	74.91	10.21	70.85	8.10	1.99	.05
Pd	59.91	12.50	60.58	11.46	.25	N.S.
Mf	57.96	10.23	51.55	9.59	2.93	.01
Pa	60.64	10.39	53.83	10.37	2.98	.005
Pt	73.87	15.73	65.05	15.32	2.58	.025
Sc	69.58	15.13	60.05	13.42	3.02	.005
Ma	56.69	9.41	56.25	9.51	.21	N.S.
Es	38.34	7.03	40.33	5.74	1.40	N.S.

Note.—Raw scores given on *?*, *L*, *F*, *K*, and *Es* (Ego-strength); T-scores on all other scales.
* *p* less than value indicated with one-tailed test for those scales where the direction was predicted; otherwise two-tailed test was used.

(and a greater number of continuers give such responses).

Hypothesis 2. The continuers are less defensive as reflected by their rejecting fewer cards (and fewer cases rejecting cards) and having a lower *F%* and *A%*.

Hypothesis 3. The continuers manifest a more persistent mental attitude by giving a greater number of total responses (*R*), and a greater number of space responses (*S*) (and a greater number of cases give *S* responses).

Hypothesis 4. The continuers more closely resemble a group of "normal" subjects when compared on the basis of *R*, *F%*, *A%*, number of cards rejected, and total number of *S*, *C*, *Y*, *V*, and *An* responses, than do the attriters.

Subjects and Procedure

The 85 patients were veterans in a mental hygiene clinic who had been diagnosed psychoneurotic without organic complications and who had been administered both the Rorschach and MMPI during the intake procedure. The patients were divided into two groups to conform to the "short" and "long" therapeutic processes suggested by Cartwright: (*a*) 40 subjects who terminated treatment prior to the 13th interview, who are referred to as the "attriters" (mean age 33.45, *SD* 7.52; mean education 10.30 years, *SD* 2.14); (*b*) 45 subjects who remained in treatment for 13 or more interviews, who are referred to as the "continuers" (mean age 33.60, *SD* 7.25; mean education 10.93 years, *SD* 2.55). The two groups do not differ significantly on mean age and education.

The 50 normal subjects were residents of a small midwestern community[2] located in the same geographical locale from which the neurotics came. They were comparable in age (mean age 35.62, *SD* 7.24) and the mean education of the neurotics suggests as well that the normals are probably similar in this respect based on average education expectancy.

All tests were individually administered. The Rorschach scoring of the original examiner was accepted for all of the normal cases as the protocols had been scored according to the Beck system (1) and checked by a second person. In the cases of the patients where the Rorschach protocols had not been scored according to Beck or where the records had been scored by psychology trainees, identification was removed from the

---

[2] The records were taken from a larger study done under the supervision of Marshall R. Jones and supported by the University of Nebraska Research Council.

Table 2

Comparison of Normal, Continuer, and Attriter Groups on Mean Scores for Certain Rorschach Variables

Rorschach variable	Normal ($N = 50$)		Continuer ($N = 45$)		Attriter ($N = 40$)	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
#*R*	24.52	12.78	24.84**	11.69	15.40**	7.72
F%	62.16	15.08	63.98	16.41	69.50	19.85
A%	49.98	14.74	46.20**	15.97	58.60**	17.18
#Rejections	.40	1.00	.49*	1.22	1.25*	1.51
#*S*	1.38	1.66	1.51*	1.60	.85*	1.24
#*C*	4.50	3.22	4.13**	3.01	1.98**	2.43
#*Y*	3.04	2.65	2.53*	2.40	1.60*	2.22
#*V*	.80	1.31	.96**	1.15	.20**	.40
#*An*	2.20	2.26	2.67**	2.71	.98**	1.65

* Continuers and attriters significantly different at .05 level (one-tailed tests).
** Continuers and attriters significantly different at .01 level (one-tailed tests).

records and they were rescored by the author. In order to eliminate some of the unreliability of scoring of the Rorschach variables being investigated, no attempt was made in stating the hypotheses to differentiate between the responses determined primarily and secondarily by the form element. There was simply a count of the responses in which the particular determinant was present.

Results

Hypothesis 1. Symptoms and affects. The results support the hypothesis that the continuers show a greater elevation of the symptom scales (*Hs*, *D*, *Pa*, *Pt*, and *Sc*) of the MMPI than the attriters. These differences were all significant at the .05 level or less except for the *Hs* scale which was in the predicted direction. The means and standard deviations of the MMPI scales for the two groups are reported in Table 1.

As predicted, the continuers gave more *C*, *Y*, *V*, and *An* responses, and a large number of cases gave responses falling in these categories. The mean differences between the two groups on the number of responses for each variable are all significant at the .05 level or less as shown in Table 2. Significantly more continuers gave *C*, *Y*, *V*, and *An* responses than did the attriters. The number and percentage of individuals in each group who gave these responses, and the chi-square tests of differences are reported in Table 3.

Hypothesis 2. Defensiveness. The continuers rejected significantly fewer cards, and significantly fewer continuers rejected cards, and gave a lower *A*%. The *F*% was lower in the continuers but not significantly so (see Tables 2 and 3).

Table 3

Comparison of Number and Percentage of Subjects in Normal, Continuer, and Attriter Groups Giving Certain Types of Rorschach Responses

Rorschach variable	Normal		Continuer		Attriter		Continuer vs. Attriter	
	No.	%	No.	%	No.	%	χ^2	*p*
Rejecting card(s)	9	(18%)	8	(18%)	22	(55%)	12.85	.001
S	31	(62%)	30	(67%)	19	(48%)	3.19	.10
C	46	(92%)	41	(91%)	28	(70%)	4.73	.05
Y	44	(88%)	34	(76%)	22	(55%)	3.98	.05
V	21	(42%)	24	(53%)	8	(20%)	10.02	.01
An	38	(76%)	33	(73%)	15	(38%)	11.06	.001

Table 4

Frequency and Assigned Weights of Prognostic Signs Occurring in the Continuer and Attriter Groups

Prognostic sign	Continuers		Attriters		χ^2	p	Weight
	Present	Absent	Present	Absent			
$Hy > Pt$	24	19	28	10	2.80	.10	1
$Hy > Sc$	29	15	33	6	3.83	.10	1
$Pd > Pa$	20	24	30	10	7.59	.01	3
$Pd > Pt$	9	35	19	20	7.39	.01	3
$Pd > Sc$	8	36	22	16	13.86	.001	4
$Ma > Pa$	15	28	20	17	2.97	.10	1
$Ma > Pt$	7	37	14	26	4.07	.05	2
$Ma > Sc$	6	38	16	22	8.42	.01	3
$R \leqq 15$	9	36	21	19	9.79	.01	3
$A\% \geqq 58$	8	37	23	17	14.42	.001	4
$Rejections \geqq 1$	8	37	22	18	12.85	.001	4
$S\% \leqq 2$	15	30	22	18	4.04	.05	2
$C\% \leqq 6$	5	40	18	22	10.48	.01	3
$Y\% \leqq 3$	12	33	18	22	3.12	.10	1
$V\% \leqq 2$	21	24	33	7	11.74	.001	4
$An\% = 0$	12	33	25	15	11.06	.001	4

Hypothesis 3. Persistence. The continuers gave significantly more S responses and a greater number of total responses (R). The difference between the number of continuers and attriters giving S responses is significant at almost the .05 level.

Hypothesis 4. The continuers were more like the normal subjects on each of the variables predicted than they were like the attriters. As was expected, the two former groups tended to differ from each other with respect to the importance of the "form" element in the responses (e.g., the normal group gave more FC and fewer CF and C responses than did the continuers, similarly with the other variables).

The continuers gave a significantly larger number of total responses than did the attriters; therefore chi squares were computed between the two groups on S, C, Y, V, and An responses corrected for R. The method of correction was that suggested by Cronbach (3, p. 411). Each variable was plotted against R for the combined groups, and the median was used as the dividing point. Then the number of cases in each group falling above and below the median were compared by chi square. The results were all significant at the .05 level or better. In addition, the same method of correction was applied to the percentage score for each variable. The chi squares were all significant at less than the .01 level except for $S\%$ and $Y\%$ which did not reach the .05 level.

In the interest of the practical significance of the variables for prognostic purposes, a cutoff point for each and for the variables combined was sought which would afford the maximum differentiation effectiveness of the measures. Therefore, an objective configural analysis, which has been described in detail elsewhere (7, 8), was applied to the MMPI profiles of the subjects and there were eight scale pairs which differentiated the two groups at the .10 level or less. Then the best cutoff point for separating the continuers and attriters on the basis of the percentage score for each Rorschach variable was determined. The resulting prognostic signs (eight MMPI scale pairs and eight Rorschach variables), frequency and significance level of occurrence of each, and arbitrary weights assigned for prognostic purposes are shown in Table 4. A combined prognostic score was determined for each patient by summing the weights of the signs present. These prognostic scores for the 85 patients were arranged in a frequency distribution and the most effective upper and lower cutoff points determined. An interval tabulation of the frequency of prognostic scores occurring in the two groups is shown in Table 5. The upper (19–43) and lower (0–10) ranges represent the most effective cutoff points for differentiating the attriters and continuers, respectively.

In previous unpublished research on the outcome of psychotherapy, the author obtained therapists' ratings of improvement on 56 of the cases used in the present study (21 attriters and 35 continuers). These cases were divided into two groups, one group containing those who showed limited or no improvement (all the attriters and 20 or 57% of the continuers) and the other group comprised of those showing moderate or extensive improvement (15 or 43% of the continuers). Hereinafter these will be referred to as the "limited improvement" and "moderate improvement" groups. A comparison of these two groups on mean MMPI profiles and Rorschach variables mentioned above reveals relationships between these scores and improvement similar to those found between these scores and continuation in treatment. The only exception was that the limited improvement group had a higher *Hs* score than did the moderate improvement group (the *Hs* was not significantly higher in the continuers than in the attriters). Patients in the moderate improvement group tended to have higher scores on *D, Mf, Pa, Pt,* and *Sc,* and to have a greater number of *C, Y, V,* and *An* responses, a larger *R,* and a lower number of rejections and *A*%. Tests of significance were not computed between the groups because of the small number of subjects in the moderately improved group and the crude rating scale used. The writer is well aware of the likelihood of contamination in the judgments of the therapists in that the ratings of improvement may have been influenced by the length of time the patients remained in treatment. A partial check on this was made by comparing the limited improvement and the moderate improvement cases of the continuer group only. The results reveal a tendency for most of the variables to be related in the predicted direction to improvement. However, these data from the unpublished research should be considered only as suggesting a relationship between the personality variables studied and improvement in psychotherapy.

Discussion

The results of this study offer support for the hypothesis that there are identifiable personality variables which are associated with the premature termination of, or continuation in, individual therapy. Also, there is evidence to suggest that all or most of these variables are related to degree of improvement as rated by therapists. Because of the very homogeneous nature of the patients used in this study, the strength and number of variables identified are probably minimal. It is expected that with diagnostically more heterogeneous groups, such as those used by Cartwright and Taylor, these differences would be more pronounced and other pertinent attributes could be identified.

The author interprets the data to indicate that those patients who stay in treatment longer, and who are likely to show greater improvement, are considerably more responsive emotionally to their perceived world than are those who drop out prematurely. They are sensitive to a wider range of affective stimuli, including both those of a more pleasurable and painful nature. The continuers are more anxious, sensitive, dependent, self-doubting, and have increased awareness of feelings of inadequacy, inferiority, and depression. They possess better potential for self-appraisal, affective reactivity and have more of a compulsive introspective attitude. The greater need of the continuers to be accepted and to receive affection is probably a motivating factor in the increased interaction with others. However, these interactions are in many ways immature. The continuers' stronger feminine interests, or less adequate identification with the cultural norms of masculinity, and greater

Table 5

Frequency and Percentage of Prognostic Scores Occurring in the Continuer and Attriter Groups and Differences Between Groups

Prognostic score	Continuers	Attriters	χ^2	p
19–43	4 (9%)	29 (73%)	33.15	<.001
11–18	16 (36%)	10 (25%)		
0–10	25 (56%)	1 (3%)	25.35	<.001

preoccupation with anatomy may be indicative of more disturbance in the sexual area. The attriters' not scoring significantly lower on the *Hs* scale as they did on *Mf* and *An* would be consistent with this view. Many clinicians consider *An* responses to be a screen for things more directly sexual in nature. There is also a greater tendency for the continuers to be more moody, to nurse grudges, and to be basically more hostile.

The repressive defense of the continuers is failing and, instead of a strengthening of this mechanism as occurs in the attriters, they are drawing upon other defenses, particularly projection, phobic and obsessive ideation, or compulsive behavior. The test pattern of the continuers is much more like that usually considered to be characteristic of patients whose primary symptoms are of a conversion or somatization nature than is the attriters'. Approximately 60% of both the attriters and continuers on whom the information was available were diagnosed anxiety reactions and about 35% of both groups were classified as conversion reactions. The continuers have an immature attitude toward life, are poorly controlled emotionally, have fear and guilt related to their sexual functioning, and pay a great deal of attention to anatomy. Their primary defense of repression, along with their tendency to convert emotional conflict into motor or sensory channels, has not been effective in alleviating the underlying anxiety; therefore, other defenses have been resorted to but to a lesser degree.

In contrast to the continuers, the attriters emphasize intellectual control and handle situations in an impersonal, matter-of-fact way. Repression has generalized to the point that they are able to acknowledge or respond to a very limited range of emotional stimulation; consequently, both their social interaction and thought processes have become very stereotyped and barren. The attriters' relative absence of emotional lability and consciously felt anxiety, plus a greater withdrawal and a more intellectual approach to their problems, suggest that their repressive and narcissistic defenses and symptoms are working sufficiently well to enable them to adjust without continued treatment, at least temporarily.

Quantitatively, the continuers and normals are very similar on the Rorschach variables investigated. However, a qualitative inspection of the data reveals that the continuers are less well integrated and manifest a more immature responsiveness.

These findings are consistent with the generally accepted indicators of better risks for therapy and with the results of the studies mentioned. There is much to suggest that feelings of anxiety, inadequacy, depression, dependency, and the potential for self-evaluation and affective responsiveness are signs of strength within the personality. While the continuers tend to have more vague somatic symptoms, an inspection of the improvement ratings reveals that there is less improvement in those patients who are extremely preoccupied with hypochondriacal complaints. The exaggerated physical concern would appear to be a defense against accepting the psychogenic etiology of their symptoms.

Summary

This study investigated the hypothesis that there are certain identifiable personality variables which are related to therapy prognosis. A group of neurotics who remained in treatment for 13 or more interviews and a group who terminated prior to the 13th interview were compared on the basis of certain Rorschach scoring categories and MMPI scales. Both of these groups were compared to a group of normal subjects on the Rorschach variables. As predicted, the continuers were less defensive, and more persistent, anxious, sensitive, and dependent than the attriters. They possessed an increased consciousness of feelings of inadequacy, inferiority, and depression and had better potential for self-appraisal, emotional responsiveness, and more of an introspective attitude. There is evidence to suggest that these personality variables are associated with improvement as well as with continuation in therapy. The continuers resembled the normals more than they did the attriters on the Rorschach measures. The results support the contention that painful affects which the neurotic experiences are signs of strength within the personality.

Prognostic scores are suggested for differentiating the continuers from the attriters.

However, the scores have not been cross-validated and should be used very cautiously.

Received May 17, 1957.

References

1. Beck, S. J. *Rorschach's test. Vol. I. Basic processes.* (2nd ed.) New York: Grune & Stratton, 1949.
2. Cartwright, D. S. Success in psychotherapy as a function of certain actuarial variables. *J. consult. Psychol.*, 1955, **19**, 357–363.
3. Cronbach, L. J. Statistical methods applied to Rorschach scores. *Psychol. Bull.*, 1949, **46**, 393–429.
4. Gallagher, J. J. Group report of a program of research in psychotherapy. In W. U. Snyder (Ed.), *Group report of a program of research in psychotherapy.* State College, Pa.: Pennsylvania State College, 1953. Pp. 21–38.
5. Mowrer, O. H. (Ed.) *Psychotherapy: Theory and research.* New York: Ronald Press, 1953.
6. Rosenberg, S. The relationship of certain personality factors to prognosis in psychotherapy. *J. clin. Psychol.*, 1954, **10**, 341–345.
7. Sullivan, P. L., & Welsh, G. S. A technique for objective configural analysis of MMPI profiles. *J. consult. Psychol.*, 1952, **16**, 383–388.
8. Taulbee, E. S., & Sisson, B. D. Configurational analysis of MMPI profiles of psychiatric groups. *J. consult. Psychol.*, 1957, **21**, 413–417.
9. Taylor, J. W. Relationship of success and length in psychotherapy. *J. consult. Psychol.*, 1956, **20**, 332.

Journal of Consulting Psychology
Vol. 22, No. 2, 1958

A Case of Folie a Deux and Projective Techniques

Edward M. Scott
Eastern Oregon State Hospital

Folie a Deux is mentioned only briefly in the literature. Psychological testing has been reported in but two articles. Another case on such an intriguing topic should be of some interest.

Recently committeed to the Eastern Oregon State Hospital were two 50-year-old, probably identical, twin brothers. They had never married and had a "spotty" work record. The hospital staff diagnosed them as paranoid schizophrenics, with folie a deux condition. "A," the dominant twin, became psychotic and influenced "B" into the same condition. A brief summary of the psychological battery is given:

On the Wechsler-Bellevue Intelligence Scale:

	"A"	"B"
Verbal Scale IQ	105	105
Performance Scale IQ	99	94
Full Scale IQ	103	100

On the Rorschach, "A" gave 13 responses, with 11 F–responses, and 11 perceptions that were sexual in nature. "B" gave 21 responses, with 7 F–responses and 7 that were sexual in content. Many of their perceptions appeared to be fundamentally of the same psychopathological nature. For example, on Card 9, "A" said, "Nasty looking. Part of a woman. Maybe the lower part"; while "B" said, "Oh! Unclean. Just the upper part. Most women are in the lower part."

Their TAT responses appeared to show similar dynamic trends. It is difficult to show a typical theme; perhaps Card 14 is significant. "A" responded, "Somebody looking out the window. Ooh! Terrible. Outside is freedom. That's a lot better. Somebody penned him up. He hurt the world. Gave it a disease." "B" responded to the same card, "Ooh! Must be waiting for somebody, all alone. Remorse, guilty, committed murder or killing. Maybe his past life caused him to do that. He'll have to pay. Probably go mad."

"A" earned a higher score, 92, than "B's" 81 on the Bender-Gestalt.

The projective tests indicated that "A" was more deteriorated yet a bit more intellectual. Using Grover's (2) and Gralnick's (1) division of folie a deux, this case appears to be one of folie impasse, although some might hold to folie communique. Both patients responded to electroshock therapy and were discharged shortly thereafter.

Psychological testing seems to be of definite value in folie a deux.

Received April 5, 1957.

References

1. Gralnick, A. Folie a deux, the psychoses of association: A review of 103 cases and the entire English literature with case presentations. *Psychiat. Quart.*, 1942, **16**, 230–263; 491–520.
2. Grover, M. M. A study of cases of folie a deux. *Amer. J. Psychiat.*, 1937, **93**, 1045–1051.

Journal of Consulting Psychology
Vol. 22, No. 2, 1958

The Prognostic Significance of Certain Behavioral Variables

A. Eskey and Ira Friedman

Cleveland Receiving Hospital and State Institute of Psychiatry

In a previous study (3) it was reported that disoriented mental patients do not improve more rapidly than oriented patients. This runs contrary to the expectation that severe symptoms reflecting personality disorganization are often more refractory than are milder overt symptoms. The authors suggest that disorientation may not be "sufficiently representative of 'acute onset . . . confusion and atypical symptoms' (1) to yield consistent results in the anticipated direction." It was suggested that other variables be investigated which might be considered more reflective of this symptom picture. The present study is an approach to this problem. Specifically, it represents an attempt to investigate the relationship of verbal activity, motor activity, and quality of thinking to length of hospitalization in order to evaluate whether these variables are related to rapidity of improvement.

Method

Two hundred randomly selected psychotic patients, male and female, between the ages of 16 and 59 were included in the study. This group exhibited a wide range of motor and verbal activity as well as varying degrees of intactness and disorganization of thought processes. The following criteria were used to exclude patients from the study: those who remained in the hospital for less than two weeks, those with any type of organic brain damage, alcoholics, mental defectives, chronic patients transferred to "custodial" institutions, those discharged as unimproved, patients leaving against medical advice, cases where language difficulties made communication unreliable, and patients who were under heavy sedation.

Each patient was rated on motor activity, verbal activity, and quality of thinking. These ratings were based upon information derived from mental status examinations obtained on each patient at the time of hospital admission. Motor and verbal activity were rated on a nine-point scale which ranged from underactivity to overactivity (two-tailed) while thinking was rated on a five-point scale ranging from intact to disorganized thinking (one-tailed). The scales are presented in Table 1.

Since one of the *E*s was the sole rater in the experiment proper, a reliability study was undertaken in order to ascertain whether other judges could reliably estimate motor activity, verbal activity, and quality of thinking from a report of a mental status examination. Ten randomly selected patients were rated by five judges, including *E*. The reliability of the average ratings computed by Ebel's (2) application of the Spearman-Brown formula to the formula for the reliability of individual ratings set forth by Snedecor (5), yields the following reliability coefficients: motor activity $r = .98$, verbal activity $r = .98$, and quality of thinking $r = .96$.

Since all the patients included in the study were discharged as improved or recovered, the prognostic significance of the variables were derived from relating them to length of hospitalization. From an operational point of view then, prognosis refers to the rapidity of improvement or recovery rather than to degree or quality of improvement. Comparisons were made between groups showing high, low, and normal verbal and motor activity, as well as those displaying intact, as opposed to distorted or disorganized, thought processes in terms of length of hospitalization. The t test,

Table 1

Scales for Verbal Activity, Motor Activity, and Quality of Thinking

Scale Value	Criterion
	Verbal Activity Scale
9	Patient is extremely circumstantial; talks almost constantly. Little or no pause in his conversation. Minimal (or no) attention paid to examiner's questions (or presence).
8	Patient is definitely overtalkative. Noticeable "push" behind speech but examiner is still largely able to direct the course of the interview.
7	Speech is noticeably more "loose" than normal. Some evidence of circumstantiality.
6	Patient is slightly overtalkative, may be mildly expansive.
5	Normal flow of speech. Patient enters freely into discussion with the examiner.
4	Speech is mildly inhibited but no real constriction is present. Speech is still largely spontaneous.
3	Patient does not communicate freely; some blocking may be noted. May answer examiner's questions reluctantly or tangentially.
2	Speech is definitely constricted but communication still takes place. Patient may be "holding back" relevant information. Little attempt to initiate or maintain conversation.
1	Minimal (or no) communication. Speech highly constricted; no pertinent information is elicited. May be mute.
	Motor Activity Scale
9	Patient is hyperactive, restless, and unable to sit still. In constant or near constant movement. May be in restraints.
8	Definite signs of excitement are present. Patient is clearly overactive but still able to exert some control over his activity.
7	Mild signs of agitation are present. Patient is more than normally active but in no respect exhibits a real loss of control.
6	Patient may be slightly restless or fidgety, e.g., tapping his foot, drumming fingers, shifting uneasily in his chair, etc.
5	Normal degree of motor activity. No evidence of retardation or agitation.
4	Patient may appear to be slightly listless or lacking in energy but is not actually retarded nor apathetic.
3	A noticeable degree of apathy or retardation is present. General motor activity is reduced but patient still exhibits some freedom of movement.
2	Physical movements are definitely retarded. Movements are slow and effortful.
1	Little or no physical movement is present. Patient is quiet or sits stiffly in chair. May be rigid.
	Quality of Thinking Scale
1	Thought processes appear to be intact. No evidence of illogical or distorted thinking.
2	Minor deviations from orderly thinking noted. No gross misinterpretation of reality is present. Some fluidity or looseness of thinking may be noted.
3	Definite evidence of disturbed thinking. Logical inconsistencies are apparent. Delusional ideas and/or hallucinations may be present. There may be some doubt that his experiences are based on reality.
4	Reality testing is obviously impaired but patient still has some contact with reality. Remnants of logical thinking are still present but in general the patient ignores or is unaware of logical inconsistencies. Delusions and/or hallucinations present and are accepted without question by the patient.
5	Patient exhibits extremely confused and bizarre thinking. Little or no appreciation of reality is present. Thoughts are so fragmented, or communication is so scant as to make more exact estimation of the degree of control impossible.

utilizing medians, was employed to calculate the significance of the differences.

Results

Each patient was rated for motor activity, verbal activity, and thinking in accordance with the descriptive statements of the scales in Table 1. The distribution of the patient population on the various scale categories is indicated in Table 2.

The patients were grouped on verbal and motor activity in terms of overactivity (high, rated 7, 8, or 9), underactivity (low, rated 1, 2, or 3), and normal activity (middle, rated 4, 5, or 6). These groups, differing in degree of activity, were compared with one another to ascertain the significance of the differences in length of hospitalization. The *extremely* active (group rated 9, and groups rated 9 and 8 combined) and the *extremely* underactive (group rated 1, and groups rated 1 and 2 combined) were also compared with one another and with the prototypic normally active group (those rated 5). The *t* test was used to test for the significance of the differences in length of hospitalization and the results are presented in Table 3.

In reference to thinking, the same procedure was followed but somewhat different groups were combined because the scale is one-tailed rather than two. In this case, differences in length of hospitalization were com-

Table 2

Frequency Distribution of Patients on the Three Scales

Scale	Scale Value									
	1	2	3	4	5	6	7	8	9	
Motor activity	8	7	13	18	66	28	19	18	23	200
Speech activity	9	25	34	7	47	9	12	35	22	200
Thinking	4	9	46	76	65	—	—	—	—	200

puted only for those groups rated 1 and 2 as compared to those rated 3, 4, and 5 combined and to those rated 5 alone. Because the *N* is so small in the combined group rated 1 and 2, a comparison was made between 1, 2, and 3 as contrasted to 4 and 5. These results are also presented in Table 3.

The results summarized in Table 3 reveal that, in the areas of motor and verbal activity, there are no significant differences in length of hospitalization. This holds true when comparing normally active with both overactive and underactive and when comparing extremes of activity with one another. An additional check was made in the area of motor activity by selecting those patients who were placed in restraints and comparing them with an equal number of randomly selected patients displaying a normal degree of motor activity. In this comparison, the two groups ($N = 38$) did not differ significantly in terms of median length of hospitalization ($t = .19$, $p = .85$). These results indicate that neither motor nor verbal activity are valid prognostic indicators of rapidity of improvement.

From Table 3 it may also be observed that significant differences are obtained when comparing patients who reveal normal or relatively normal thinking with those who display distorted or disorganized thinking. These differences are not in the anticipated direction, however, and it appears that those patients who have little or no disturbance of thought processes are more likely to be discharged sooner. Unfortunately, the number of cases displaying adequacy of thought processes is small ($N = 13$) so that the generality of conclusion is limited. When the group rated three on thinking, i.e., somewhat pathological, was included along with those with more normal thinking processes and compared with patients displaying greater thought pathology, the results failed to reach acceptable levels of significance. However, even here, a

Table 3

Significance of Differences in Length of Hospitalization Between Groups Differing in Scale Ratings

Scale Ratings of Comparison Groups	Verbal Activity		Motor Activity		Quality of Thinking	
	t	*p*	*t*	*p*	*t*	*p*
123–789	.46	.64	.97	.33	—	—
123–456	.40	.70	.56	.58	—	—
456–789	.55	.58	.62	.54	—	—
12–89	.23	.82	1.27	.20	—	—
12–5	.55	.58	.46	.65	2.15	.04
5–89	.36	.72	.34	.73	—	—
1–9	.20	.84	.10	.92	—	—
1–5	.44	.64	.04	.97	—	—
5–9	.29	.77	.45	.66	—	—
12–345	—	—	—	—	2.39	.02
123–45	—	—	—	—	1.43	.15

tendency for more rapid discharge of patients whose thinking was more intact upon admission was noted.

As a final check on the data, the upper quarter of each group was compared with the lower quarter by means of the Mann-Whitney U test (4). In this comparison, only thinking proved significant ($p = .02$), whereas motor activity ($p = .33$) and verbal activity ($p = .23$) proved nonsignificant as in the previous tests. This nonparametric technique was used not only because of nonnormal distribution of the data, but also to give more appropriate weight to extreme scores which tend to be underweighted with medians and overweighted using means.

Discussion

Overactivity and underactivity in verbal and motor behavior and a greater disorganization of thought processes do not appear to be associated with more rapid improvement as measured by length of hospitalization. In the area of thinking, as a matter of fact, the less the disorganization of thought processes at the time of admission the greater is the tendency for more rapid discharge. These findings shed doubt upon the generalization that more extreme symptomatology responds more rapidly to treatment.

In a previous article (3), the authors related the impression that severe symptomatology is more refractory than milder overt symptomatology to Gestalt memory principles of similarity and contrast, i.e., that we tend to recall those patients whose final state differs markedly from their condition upon admission. The present study suggests that improvement, sufficient to warrant discharge, is nondifferential between patients displaying extreme or mild symptomatology *in temporal terms*. It is important to point out clearly, however, that no attempts were made to examine the *qualitative* degree of improvement. Since all patients were discharged as improved, we might hypothesize that attempts to measure qualitative changes between condition upon admission and at discharge would yield greater change in those patients who initially displayed more deviant behavior. We do not know, however, whether the discharge condition of patients differing upon admission in terms of severity of symptomatology would yield differences in terms of some absolute criteria of "normality." It appears that the evaluation of the degree and nature of qualitative change, and the comparison of discharge condition to some absolute criteria of "normality" is a fruitful area for further research to shed more light on this problem. Replication of the present study and further studies investigating other symptom pictures are also necessary.

Summary

Two hundred psychotic patients discharged as improved were rated on verbal activity, motor activity, and quality of thinking displayed at the time of hospital admission. Comparisons were made to determine whether differences in degree of overt symptomatology were associated with rapidity of improvement as measured by length of hospitalization. The results indicate that degree of activity is not associated with rapidity of improvement but that patients with more intact thought processes at the time of admission tend to be discharged more rapidly. The findings shed doubt upon the generalization that more extreme symptomatology responds more rapidly to treatment. Differences in viewing improvement in temporal and qualitative terms were discussed in reference to suggestions for further research.

Received May 6, 1957.

References

1. Chase, L. S., & Silverman, S. Prognostic criteria in schizophrenia. *Amer. J. Psychiat.*, 1941–42, **98**, 360–368.
2. Ebel, R. L. Estimation of the reliability of ratings. *Psychometrika*, 1951, **16**, 411.
3. Eskey, A., Friedman, Gladys M., & Friedman, I. Disorientation as a prognostic criterion. *J. consult. Psychol.*, 1957, **21**, 149–151.
4. Mann, H. B., & Whitney, D. R. On a test of whether one of two random variables is statistically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.
5. Snedecor, G. *Statistical methods*. (4th ed.) Ames, Iowa: Iowa State Coll. Press, 1946.

Journal of Consulting Psychology
Vol. 22, No. 2, 1958

Self Acceptance and Marital Happiness[1]

Daniel Eastman
Teachers College, Columbia University

Previous researches on the personality correlates of marital happiness or adjustment have tended to rely on heterogenous and eclectic personality measures of limited theoretical and clinical significance (2, 3, 6, 7, 12). In the present study, self acceptance was chosen as a personality measure for investigation because it is a relatively homogenous variable based on a developed theory of personality with explicit clinical implications (8, 9, 10, 11, 15).

Reduced to its simplest terms, the experiment consisted in scoring a sample of married couples on two questionnaires, one for marital happiness (MH) and one for self acceptance, and correlating the two sets of scores. This simple scheme was complicated by further investigating: (*a*) the relation of MH to the additional personality variables, acceptance of others and psychological status; (*b*) the relation of MH to all the personality variables in the subjects' mates as well as in the subjects; (*c*) the relation of MH to the mates' MH, that is, the role of mutual happiness; and finally, (*d*) the effect of sex differences on the correlations between MH and the personality variables.

Method

Marital happiness was measured with the instrument constructed by Wallace (14), in essence a compilation of the most discriminating items in the Terman (12) and Burgess and Cottrell (2) questionnaires.

Self acceptance (s) was measured as the discrepancy between *S*s' ratings of self and self ideal, using an abbreviated form of the rating scales developed by Bills, Vance, and McLean (1). Acceptance of men (m) and of women (w) was measured on the same scales as the discrepancy between ratings of men and women in general, and self ideal; acceptance of others (o) being then taken as the mean of acceptance of men plus acceptance of women $(m/2 + w/2)$. The variable called psychological status was calculated as the signed difference between acceptance of men, women, or others, and self acceptance (m-s, w-s, or o-s). Psychological status corresponded to what are sometimes called feelings of superiority or inferiority.

*S*s were gathered in groups of five or more couples brought together by colleagues ostensibly for social purposes. All couples in every group agreed to fill out the questionnaires. In order to encourage frankness, the questionnaires were anonymous, and couples were separated before filling them out. Each couple was required to put its completed questionnaires in the same manila envelope before returning them, so that the investigator was able to determine married pairs, while preserving anonymity and interspouse secrecy. When some eight spoiled blanks were eliminated, the sample consisted of 50 couples married two years or more, and 14 couples married less than two years.

Results

Table 1 shows the means and standard deviations of all the variables for the sample of 50 couples married more than two years. The

---

[1] This paper is based upon a dissertation submitted to the Department of Family Life, Teachers College, Columbia University, in partial fulfilment of the requirements for the degree of doctor of philosophy; committee members, Ernest G. Osborne, chairman; Arthur T. Jersild and Laurance F. Shaffer. The original report (4) may be consulted for more detailed treatment of the rationale, methods, and results.

Table 1

Marital Happiness and All Variables, by Sexes, for 50 Couples Married More Than Two Years (Means and standard deviations)

	Husbands ($N = 50$)		Wives ($N = 50$)	
Variable	Mean	*SD*	Mean	*SD*
Marital happiness	108.4	25.49	112.6	27.15
Self acceptance	36.2	15.97	36.4	14.66
Acceptance of men	46.8	10.97	45.0	14.90
Accept. of women	51.7	10.00	51.1	16.50
Status re men	10.6	15.90	8.6	17.60
Status re women	15.5	14.53	14.7	17.69

split-half reliabilities corrected by the Spearman-Brown formula were: for self acceptance, .89; for acceptance of men, .82; for acceptance of women, .90. Wallace reported a corrected reliability of .90 for his test of marital happiness.

The means for all the variables reported here were not significantly different from those obtained in other investigations with the same questionnaires in clinically "normal" groups (4).

The chief point to be noted in Table 1 is the complete lack of any significant difference in means between the husbands and wives in the sample on any of the variables measured. This lack of mean difference between sex groups was confirmed in another sample not reported here (4). Under these circumstances, any observed difference in the marital performance of the two sex groups could not be explained on the basis of different average endowments of the two sexes.

It may be noted that both husbands and wives scored women in general significantly less acceptable than men in general, a bias also confirmed for both sex groups in another sample not reported here (4). Statistical tests in the further course of the investigation showed this bias to have no significant correlation with marital happiness in either husbands or wives or their mates.

Marital Happiness and Self Acceptance

Table 2 shows all the zero-order correlations between MH and self acceptance. MH was significantly related to self acceptance both in the *S*s themselves and in their mates. If sex groupings were combined, the correlation between MH and the *S*s' own self acceptance was significant at the .0005 level and between MH and mates' self acceptance at the .0001 level.

The correlation between the husbands' MH and their wives' self acceptance was the highest obtained for self acceptance in a single mate. This higher correlation was probably not the result of accident. In the sample of 14 couples married less than two years, the correlation between husbands' MH and wives' self acceptance was .53, the only significant correlation with self acceptance obtained in that small sample.

The correlation of MH with the summed self acceptance of both mates was considerably higher than the correlation with either mates' self acceptance alone, in part because the mates' self acceptance scores were largely independent of each other ($r = .24$). About twenty-five per cent of the variance in MH was accounted for by the summed self acceptance of both mates; a figure that compares favorably with the percentage of MH accounted for by any other single variable, or any other combination of variables, previously investigated. It should be pointed out, however, that the variables employed in previous investigations were not summed for both mates, since the investigators either did not, or could not, obtain a significant correlation between MH and mates' scores on the variables employed.

Table 2

Correlation of Marital Happiness with Subjects' and Mates' Self Acceptance, by Sexes, for 50 Couples Married More Than Two Years

	Husbands ($N = 50$)		Wives ($N = 50$)	
Marital Happiness Correlated with—	r	p	r	p
Self acceptance	.37	.005	.35	.01
Mates' self acceptance	.47	.0005	.35	.01
Self acceptance of both mates	.53	.0005	.44	.005

Table 3

Correlation of Marital Happiness with Other Variables, by Sexes, for 50 Couples Married More Than Two Years

Marital Happiness Correlated with—	Husbands ($N = 50$)		Wives ($N = 50$)	
	r	p	r	p
Acceptance of				
Men	.01	—	.26	.05
Women	.12	—	.32	.025
Others	.07	—	.32	.025
Status re				
Men	.36	.005	.07	—
Women	.32	.025	−.01	—
Others	.36	.005	.05	—

Note.—The score for acceptance of others was the mean of the scores for acceptance of men plus acceptance of women.

Marital Happiness and the Other Variables

Table 3 shows the correlations between MH and the *S*s' own scores for acceptance of others and for psychological status; Table 4 shows the correlations between MH and the mates' scores for acceptance of others and psychological status.

It will be seen that the MH of both husbands and wives correlated significantly with the wives' acceptance of others, but not with the husbands' acceptance of others; and that the MH of both husbands and wives correlated significantly with the husbands' psychological status but not with the wives' psychological status. Acceptance of others was maritally operative only in the wives, psychological status only in the husbands.

In view of the double check afforded by separate scores for acceptance of men (m) and acceptance of women (w), and of the further double check afforded by cross validation between *S*s and mates, it seems fairly safe to suppose that the sex pattern revealed in these correlations exists in the population of couples married more than two years.

Marital Happiness and Mutual Happiness

The correlation between the MH scores of the husbands and their wives was .74 for the 50 couples in the present sample. This figure appears to indicate a considerable dependence of MH on mates' marital happiness, that is, a high degree of mutuality in MH.

However, the figure .74 was unreliable because the MH test itself is partly composed of items that deal with objective facts, such as the amount of disagreement between the mates. The correlation may therefore have represented *in part* nothing more than the tendency of the mates to see objective facts about their marriage in the same way. Nevertheless the positive correlation between the MH scores of the mates obtained in this study, and in every other study using similar instruments, probably does indicate some real interdependence of MH, that is, real mutuality, as Terman supposes (12, pp. 80–83).

In Table 5, partial correlations are shown between MH and self acceptance where mates' MH was held constant, that is, where all the mutuality, whatever its unknown size, was eliminated statistically. It may be seen in the first row of the table that under these conditions MH showed no significant correlation with *S*s' own self acceptance in either sex group, in contrast to the highly significant correlations obtained earlier when mutuality was not eliminated.

Apparently, *S*s' own self acceptance is related only to that portion of MH which is

Table 4

Correlation of Marital Happiness with Mates' Scores on Other Variables, by Sexes, for 50 Couples Married More Than Two Years

Marital Happiness Correlated with—	Husbands ($N = 50$)		Wives ($N = 50$)	
	r	p	r	p
Mates' accept. of				
Men	.31	.025	.08	—
Women	.28	.025	.03	—
Others	.31	.025	.06	—
Mates' status re				
Men	.13	—	.30	.025
Women	.13	—	.37	.005
Others	.13	—	.35	.01

Note.—The score for acceptance of others was the mean of the scores for acceptance of men plus acceptance of women.

Table 5

Partial Correlation of Marital Happiness with Self Acceptance, by Sexes, Holding Mates' Happiness Constant, for 50 Couples Married More Than Two Years

Marital Happiness Correlated with—	Husbands ($N = 50$)		Wives ($N = 50$)	
	r	p	r	p
Self acceptance	.18	—	.00	—
Mates' self acceptance	.33	.01	.13	—
Mates' self acceptance[a]	.32	.025	−.10	—

[a] Indicates second-order partial correlation holding *Ss'* self acceptance constant as well as mates' marital happiness.

mutual; and is probably highly related to it, if it could be separated from MH in general.

In the second row of Table 5 it will be seen that the husbands' MH continued to have a significant correlation with their wives' self acceptance, regardless of the wives' own happiness in the marriage; and in the third row the same relation held even when the husband's own self acceptance was also held constant. Some unknown factor in the wives, not present in the husbands, caused the wives to make their husbands happier, regardless of their own happiness.

Marital Happiness and Sex Differences

The data bearing on the relation of MH to sex differences have now all been presented, and may be summarized as follows: (*a*) the husbands' MH showed a markedly higher zero-order correlation with their wives' self acceptance; (*b*) this correlation remained uniquely significant even when the wives' own happiness was held constant by partial correlation; (*c*) both mates' MH correlated with acceptance of others in the wives only; (*d*) both mates' MH correlated with psychological status in the husbands only; and (*e*) the MH of both husbands and wives was unaffected by the tendency of both sexes to rate women in general less acceptable than men.

In short, the wives were more influential in marriage, more persistent, less concerned with interpersonal status, and indifferent to sexual inferiority. There is considerable probability that this situation obtains among the population of wives married more than two years.

The Newlywed Sample

The data for the sample of 14 couples married less than two years are not presented in detail. It was assumed that the relatively unsettled state of the criterion, marital happiness, would tend to depress the correlations, and this proved to be the case in general. For instance, the interspouse correlation for MH, which was .74 in the older marriages, was only a nonsignificant .39 in the newer marriages. There was evidence of a somewhat different pattern of correlations in the new marriages; for instance, the husbands appeared to be more dependent on their wives' personality traits than the wives on their husbands' personality traits. But the sample was too small, and too few of the correlations significant, to merit further description of the results.

Discussion

The question naturally arises whether the correlation between marital happiness and self acceptance obtained in this study indicates that self acceptance causes MH, or MH causes self acceptance. This is a moot question. Most theoreticians of the self believe self acceptance to be the outcome, in large measure, of childhood experiences. If this is the case, then the results in the present research would seem highly compatible with Terman's finding, supported by every subsequent investigation, that childhood background is one of the most discriminating factors for marital happiness (12, p. 372). Self acceptance may be the variable intervening in time between childhood and marriage.

In order to make a conclusive test of the causal relation, however, it would be necessary to obtain self acceptance scores before the marriages, which was not done. In support of the causal effect of personality in general the studies of Kelly (4), Burgess and Wallin (3, p. 536), and Terman and Oden (13, pp. 257–260) may be cited, where in each instance a personality measure before marriage, or divorce, correlated with MH after marriage, or with divorce as a criterion.

The finding in this study that wives continue to influence their husbands' MH regardless of their own MH is compatible with the conclusions of Burgess and Cottrell (2, pp. 341–349), who cite it as one of their six main conclusions that American wives make the major adjustment in marriage. Making adjustments would seem to imply greater concern with and influence on the mate, since it is the mate that is adjusted to, in part.

The failure of MH to correlate with the positive social trait, acceptance of others in the husbands was unexpected, but probably accounted for by the fact that MH did correlate with the husbands' psychological status. Psychological status included acceptance of others, from which it was derived mathematically. Apparently, some third factor in husbands, perhaps competitiveness, makes status more important to them than to their wives.

Summary

With a sample of 50 couples married more than two years it was shown that:

Marital happiness is related to self acceptance, acceptance of others, and psychological status in both subjects and their mates; to self acceptance in both sexes, to acceptance of others probably only in wives, and to psychological status probably only in husbands.

The relation of marital happiness to self acceptance, acceptance of others, and psychological status is affected in several other measurable ways by average psychological differences between the two sexes.

Received April 24, 1957.

References

1. Bills, R. E., Vance, E. L., & McLean, O. S. An index of adjustment and values. *J. consult. Psychol.*, 1951, **15**, 256–261.
2. Burgess, E. W., & Cottrell, L. S. *Predicting success or failure in marriage.* New York: Prentice-Hall, 1939.
3. Burgess, E. W., & Wallin, P. *Engagement and marriage.* Philadelphia: Lippincott, 1953.
4. Eastman, D. Self acceptance and marital happiness. Unpublished doctoral dissertation, Teachers Coll., Columbia Univer., 1956.
5. Kelly, E. L. Concerning the validity of Terman's weights for predicting marital happiness. *Psychol. Bull.*, 1939, **36**, 202–203.
6. Kelly, E. L. Marital compatibility as related to personality traits of husbands and wives as rated by self and spouse. *J. soc. Psychol.*, 1941, **13**, 193–198.
7. Locke, H. J. *Predicting adjustment in marriage.* New York: Holt, 1951.
8. McDougall, W. *An introduction to social psychology.* (30th ed.) London: Methuen, 1950.
9. Muller-Freienfels, R. *The evolution of modern psychology.* New Haven: Yale Univer. Press, 1935.
10. Rogers, C. R., & Dymond, Rosalind F. *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954.
11. Symonds, P. M. *The ego and the self.* New York: Appleton-Century-Crofts, 1951.
12. Terman, L. M. *Psychological factors in marital happiness.* New York: McGraw-Hill, 1938.
13. Terman, L. M., & Oden, Melitta H. *The gifted child grows up.* Vol. IV. *Genetic studies of genius series.* Stanford, Cal.: Stanford Univer. Press, 1947.
14. Wallace, K. Construction and validation of marital adjustment and prediction scales. Unpublished doctoral dissertation. Univer. of Southern California, 1947.
15. Zuckerman, M., Baer, M., & Monashkin, I. Acceptance of self, parents and people in patients and normals. *J. clin. Psychol.*, 1956, **12**, 327–332.

Bureau Ednl. Psy. Research
DAVID HA AINING COLLEGE
Dated
Accs. No

Journal of Consulting Psychology
Vol. 22, No. 2, 1958

The Social Desirability Factor in Edwards' PPS[1]

Daniel Kelleher
University of Washington

Edwards has published a forced choice self-assessment inventory, Personality Preference Schedule (PPS), designed to give scores on personality variables that are free from the effects of social desirability (1). The inventory consists of 225 item pairs, each pair made up of two items rated by college student judges to have similar social desirability values. It has been shown that correlations between scores on the 15 variables and social desirability are low (1). The present study was designed to see whether social desirability still operated on choice of item within item-pairs, however.

One hundred and one males and 101 females from beginning psychology courses were given the PPS and the Social Desirability (SD) scale. The latter scale, described by Edwards (1), yields scores measuring the extent to which the subjects endorse statements in a socially desirable way. Point-biserial correlations were computed separately for each sex between SD scale score and choice of A or B for each of the 210 different item pairs. These correlations were found to cluster around zero, with only slightly greater than chance occurrence of significant correlations.

There are 48 item-pairs in which the difference between judged social desirability rating of the items in the pair is relatively great. Correlations were computed so that if social desirability was operating in item choice, this group of 48 pairs would show preponderantly plus correlations. No such effect was found.

A special score, HSD score, was computed by counting the number of times the subject chose the item in a pair with the higher social desirability rating, no matter how small the difference (all 210 pairs used). The mean HSD score for both males and females was higher than that expected by chance at the .01 level of confidence. Correlations between SD scale scores and HSD scores for each of the sexes showed insignificant correlations for both.

On the basis of the above results, it was felt social desirability played an insignificant role in item responses on the PPS.

Brief Report.
Received November 13, 1957.

Reference

1. Edwards, A. L. *Manual, Edwards Personal Preference Schedule*. New York: Psychological Corp., 1954.

---

[1] An extended report of this study may be obtained without charge from Daniel Kelleher, 1201 Campus Parkway, Seattle, Wash., or for a fee from the American Documentation Institute. Order Document No. 5469, remitting $1.75 for microfilm or $2.50 for photocopies.

Journal of Consulting Psychology
Vol. 22, No. 2, 1958

Six Measures of Self-Concept Discrepancy and Instability: Their Interrelations, Reliability, and Relations to Other Personality Measures[1]

Gene Marshall Smith
Massachusetts General Hospital

During recent years, largely as a result of the emphasis placed on the self concept by Rogers (19, 20), many experimental examinations of self-regarding attitudes have been reported (e.g., 1, 2, 3, 7, 18). One measure which has frequently been studied is the discrepancy between the "Self" and the "Ideal Self," i.e., the discrepancy between what the individual says he is and what he says he would like to be. The present paper deals with this measure and five new measures of self concept. Three of the five new measures deal with the instability, over time, of the concepts of "Self," "Ideal Self," and "Social Self." In this report, the concept of "Social Self" refers to the individual's notion of how he is seen by other people. The remaining two measures deal with the discrepancy, at a given point in time, between "Ideal Self" and "Social Self" and the discrepancy between "Self" and "Social Self."

The present investigation was not designed to test specific hypotheses involving the six self-concept measures mentioned above. Rather, it was designed to determine the interrelations among these six measures, to evaluate their reliability, and to explore the relations between them and certain other measures of personality. The six self-concept measures are given operational meaning in the method section and are given psychological meaning in the discussion section.

Method

Self-Concept Test

Twenty-four college men were given a self-concept test consisting of 29 personality descriptive phrases. An attempt was made to cover a broad range of needs, attitudes, and behavior tendencies with the 29 phrases. Some phrases were relevant to certain of Cattell's 16 personality factors (5); some were relevant to certain of Edwards' variables (9); some were relevant to certain of the needs listed by Murray (17). Others were relevant to no particular list of personality characteristics. Examples of the 29 traits are: (*a*) tends to express anger easily and is apt to be found in arguments; (*b*) tends to be energetic, alert, spirited, and enterprising; has great vitality and drive; (*c*) tends to think of people as being somewhat selfish, unfair, and self-seeking; (*d*) tends to be cooperative, agreeable, and ready to meet people at least halfway.[2]

Administration. Each phrase was typed on a 3 × 5 card. The experimenter gave the *S* the 29 cards, one at a time, and asked him to sort them into five piles on the basis of the degree to which they described his "Ideal Self." No restrictions were imposed concerning the number of cards which could be placed in a given pile. After this first sorting, the *S* re-sorted these same 29 cards into five piles on the basis of the degree to which they described his "Self." Finally, the *S* re-sorted them on the basis of the degree to which they described his "Social Self." "Ideal Self" was defined for the *S* as the self he would like to be—not what he thinks he ought to be but what he would like to be. "Self" was defined in terms of what the *S* thinks he actually is. "Social Self" was defined in terms of the *S*'s notion of what other people think of him. The meaning of other people was specified for the *S* by instructing him as follows: "By other people we . . . are referring to people who know you well enough to hold opinions about you, but people who are not your five or six closest friends."

---

[1] This report is based on work carried out during 1955 and 1956, supported by grants to Henry K. Beecher, from the Medical Research and Development Board of the Department of the Army and the United States Public Health Service. A pilot study investigating the six self-concept variables, carried out in 1953 at the University of Rochester, was supported by a grant to G. R. Wendt, from the Office of Naval Research. The author wishes to express his appreciation to R. L. Burke and J. M. Hoffman for their assistance with the collection and statistical examination of the data, and to those colleagues whose critical reading greatly improved the manuscript: E. L. Cowen, R. B. Cattell, W. Finegar, D. Briggs, and A. Couch.

[2] To save printing costs, the complete list of 29 traits is not presented here. It has been deposited with the American Documentation Institute. Order Document No. 5435, remitting $1.25 for microfilm or $1.25 for photocopies.

The five piles into which the *S* sorted the 29 traits were labeled in the same manner for all three trait sortings. The labels used were "Almost All of the Time," "Usually," "About Half of the Time," "Seldom," and "Almost Never." When a trait was put in Pile 1, it was given a score of one. When it was put in Pile 2, it was given a score of two, etc. Thus, if a given *S* placed a given card in Pile 1 when sorting on the basis of "Ideal Self," Pile 3 when sorting on the basis of "Self," and Pile 2 when sorting on the basis of "Social Self," it was interpreted as meaning he would like that trait to be characteristic of him almost all of the time; he thinks that trait is characteristic of him about half of the time; and he thinks other people would say that trait is usually characteristic of him.

Testing occasions. The *S* took the complete self-concept test on three separate occasions, once when medicated with morphine, once with amphetamine, and once with placebo. The three medications were administered subcutaneously in doses that varied according to body weight. The *S*s received, per 70 kg. of body weight, 10 mg. of morphine, 15 mg. of amphetamine, and 1 ml. of physiologic saline solution as a placebo. The three testing occasions were separated by intervals of approximately one week. Order of drug administration was balanced.

The three discrepancy measures (discrepancy between "Self" and "Ideal Self," discrepancy between "Social Self" and "Ideal Self," and discrepancy between "Self" and "Social Self") were based on the ratings obtained under the placebo condition only. Thus these measures were in no way influenced by the pharmacologically active ingredients in morphine and amphetamine. The three instability measures (instability of "Self," "Ideal Self," and "Social Self") were based on the ratings obtained under all three conditions of medication. However, since the self-concept test was only one of a series of tests administered to evaluate drug effect, it was given approximately 4½ hours after medication; and at this time, the remaining drug effects were weak (12, 21). Thus, at present we believe the point to be emphasized concerning the instability variables is not that they reflect instability produced by drugs but, rather, that they reflect instability occurring over time. This belief is now being checked in a separate sample of *S*s tested on three separate nonmedicated occasions.

Three discrepancy measures. The following method was used to determine each *S*'s discrepancy between "Self" and "Ideal Self." The unsigned difference between the rating an *S* gave a trait on the "Self" dimension and the rating he gave that trait on the "Ideal Self" dimension was taken as the measure of his discrepancy between "Self" and "Ideal Self" with respect to that particular trait. The absolute difference between an *S*'s "Self" rating and his "Ideal Self" rating was obtained for each of the 29 traits, and the mean of his 29 absolute differences was taken as the measure of his discrepancy between "Self" and "Ideal Self" for the test as a whole. The discrepancy between "Self" and "Social Self" and the discrepancy between "Ideal Self" and "Social Self" were obtained in like manner.

The mean of discrepancies method was employed rather than the more frequently used correlation method because it was felt that the former is less influenced by such factors as number of traits comprising the trait-pool, response set bias, etc., than is the latter.

Three instability measures. Each *S*'s "instability of Self" score was obtained as follows. The variance of the three ratings a particular *S* gave a particular trait on the "Self" dimension on the three test occasions was taken as that *S*'s degree of "instability of Self" for that particular trait. The variance of each *S*'s three ratings was obtained for each of the 29 traits. The mean of an *S*'s 29 variances was taken as the measure of his "instability of Self" for the test as a whole. Measures of "instability of Social Self" and "instability of Ideal Self" were obtained in a like manner.

Non-Self-Concept Personality Measures

In addition to the six self-concept measures, 49 "non-self-concept" personality measures were obtained for each *S*. Thirty-six of these were obtained from the following personality inventories: The Cattell 16 Personality Factors Questionnaire (4); the Maslow Security-Insecurity Inventory (16); the Manson Evaluation (15); the Edwards Personal Preference Schedule (8); the Minnesota Thinking, Social and Emotional Introversion-Extraversion Test (10). (The Edwards Test was taken under placebo; the rest were taken during the five nonmedication testing sessions mentioned in the following paragraph.) Of the 13 "non-self-concept" measures which were not obtained from personality inventories, 12 were measures of average-mood and one was a measure of self-insight.

Average-mood measures. Information concerning average-mood was obtained as follows: A total of 15 mood samples was obtained for a given *S* by having him rate himself on a seven-point scale on each of 12 bipolar mood items. These 15 samples of mood were obtained on the placebo day and on five nonmedication days, two of which preceded and three of which followed the three days on which medications were given. The mood items were presented to the *S*s in the form of bipolar seven-point scales, where each of the two poles was defined by a brief sentence. For example, the extremes of one mood item were specified in terms of the following two sentences: "I feel a little grouchy and probably would get mad easily." "I feel very good natured and probably would not get mad easily." The following is a listing of the remaining 11 items (abbreviated): sleepy vs. awake; uneasy vs. secure; mind alert vs. mind slow; dreamy vs. not dreamy; happy vs. sad; clear-headed vs. groggy; confident vs. shy; peppy vs. no pep; worried vs. not worried; calm vs. restless; unsociable vs. sociable.

The mean of a given *S*'s 15 ratings for a given mood item was taken as the measure of his average-mood with respect to that particular item. The 12 items comprising the mood questionnaire are based on a set of mood items developed by Lasagna, vonFelsinger, and Beecher (13).

Self-insight measure. The self-insight measure was obtained as follows: Two psychologists, who had interviewed and tested each *S* for about 20 hours, rated each *S* on a five-point scale on each of the 29 personality traits comprising the self-concept test. After rating a given *S* on a given trait, the psychologist evaluated his confidence in that rating, marking it high, medium, or low. The mean of the two psychologists' ratings of an *S* on a given trait was then compared with the *S*'s own rating on that trait on the "Self" dimension under placebo. The amount of agreement between the rating the *S* gave himself on that trait and the mean rating given to him by the two psychologists was taken as a measure of his self-insight with respect to that particular trait. The discrepancy between a given *S*'s "Self" rating and the mean rating given to him by the two psychologists was obtained for each of the 29 traits. However, not all of the 29 discrepancies contributed to a given *S*'s self-insight score. One of the following two criteria had to be met before a trait was used: (*a*) a trait was used if both psychologists had rated the *S* exactly the same way on that trait *and* each psychologist had recorded high or medium confidence in his rating; (*b*) a trait was used if the two psychologists had disagreed by only one scale point *and* both had recorded high confidence in their rating of that *S* on that trait. The average number of traits meeting one or both of the above criteria was 18.5 per *S*.

Results

Intercorrelation Among the Six Self-Concept Measures

Table 1 shows that each of the six variables is positively correlated with every other variable and that 11 of the $n(n-1)/2 = 15$ dif-

Table 1

Correlations Among the Six Self-Concept Variables
(N = 24)

	Discrepancy Variables		Instability Variables		
Variable	Self–Social	Ideal–Social	Inst. Self	Inst. Ideal	Inst. Social
Discrepancy					
Self–Ideal	+.81**	+.74**	+.56**	+.28	+.71**
Self–Social		+.51*	+.71**	+.21	+.65**
Social–Ideal			+.45*	+.31	+.47*
Instability					
Self				+.50*	+.68**
Ideal					+.33

* Significant at .05 level by two-tailed test.
** Significant at .01 level by two-tailed test.

ferent comparisons are significant at or beyond the .05 level, using a two-tailed test. The correlation coefficients were obtained using the formula for the Pearson product–moment r. Each of the six variables was distributed approximately normally.

Search for Artifacts

Finding everything correlated with everything is rather unusual. Therefore, before accepting these correlations at face value, the procedure used to obtain the self-concept scores and the correlations among them was examined in an effort to discover possible artifacts. For example, the possibility that the positive correlations were due to some sort of common response set was examined. In searching for evidence relevant to this possibility, it was found that *S*s differed from one another with respect to the way in which they used the five-point scale. For example, some *S*s used predominantly neutral points on the scale, whereas others used predominantly extreme points on the scale. Similarly, some *S*s tended to use only a limited range of the scale, whereas others tended to use the full range. If either tendency had been found to be correlated with any of the self-concept scores, the intercorrelation among the six self-concept variables would have become suspect. This would have been particularly true if the tendency to use neutral points or the tendency to use a restricted range had been found to be associated with low discrepancy or low instability. However, neither of these tendencies was found to be significantly correlated with any of the six self-concept variables.

In searching for possible artifactual meanings of the correlations among the self-concept variables, the question arose as to whether or not certain *S*s got higher instability and discrepancy scores than other *S*s simply because they were less careful and conscientious in making their responses. If amount of instability and discrepancy were correlated with amount of carelessness in taking the test, the correlations among the six self-concept variables would be spuriously increased. The Edwards consistency variable was taken as a measure of carefulness-carelessness in "personality-test-taking." (An *S*'s score on the Edwards consistency variable depends on the number of times he responds to identical test items in the same way.) The fact that this variable was not found to be significantly correlated with any of the six self-concept variables was taken as evidence against the assumption that correlations among the six self-concept variables were spuriously increased by variation among the *S*s with respect to carefulness-carelessness in taking the self-concept test.

Another possible source of error which was investigated was the fact that certain of the six self-concept scores are not based on completely independent data. For example, the discrepancy between "Self" and "Ideal Self" and the discrepancy between "Social Self" and "Ideal Self" contain a common element: "Ideal Self." This problem exists in the discrepancy vs. discrepancy correlations but not in the instability vs. instability correlations and only to a very slight degree[3] in the instability vs. discrepancy correlations. Table 1 indicates that although the correlations of the first of these three types are somewhat higher than those of the second and third types, the latter two types are nonetheless consistently high (8 out of 12 being significant at or beyond the .05 level using a two-tailed test).

It will be remembered that instability scores are based on variance values. An examination of the nature of the distributions of the variances indicated that they are distributed logarithmically rather than normally. Because the variances are logarithmically distributed, extreme variance values on two or three of the 29 traits could greatly increase an *S*'s total instability score. It seemed possible that in some way this might be spuriously increasing the correlations among the three instability variables. To determine whether or not the instability scores were indeed being unduly influenced by a few items with high variance values, a new method of determining instability of "Self," "Social Self," and "Ideal Self" was developed which avoided this problem, and correlations between scores yielded by the two methods were determined. With the new method, an *S*'s instability score depended on the number of traits, out of a possible 29, to which that *S* gave the same rating on all three test occasions. This new method of obtaining instability scores gave the *S*s scores which were very similar to those obtained with the mean of variances technique. The correlations between instability scores based on the two methods were + .86, + .84, and + .79 for instability of "Self," "Social Self," and "Ideal Self," respectively.

Perhaps the most convincing evidence against the supposition that the intercorrelation among the six self-concept variables is due to artifacts was obtained by comparing the discrepancy vs. discrepancy correlations and instability vs. instability correlations with the discrepancy vs. instability correlations. As was pointed out before, the discrepancy scores are based on the subtractive relationship between ratings of the traits with respect to two different "dimensions" on the same occasion (the placebo day), whereas the instability scores are based on the variance of ratings of the traits with respect to the same "dimension" over three different occasions. That is, the discrepancy scores and instability scores were obtained using different operations and (with the minor exception mentioned in Footnote 3) were based on different data. In view of this, it seems reasonable to assume that sources of error not discovered in the search for artifacts, discussed above, would be much more likely to affect the *within* class correlations (discrepancy vs. discrepancy, or instability vs. instability) than the *across* class correlations (discrepancy vs. instability). Following this assumption, all 15 r values were transformed to z values, the mean z value for the six *within* class correlations and the mean z value for the nine *across* class correlations were obtained, and these two mean z values were transformed back to r values. The corrected average r value for the across class correlations was high (.51) and was encouragingly close to the corrected average r value for the within class correlations (.62).

Reliability of the Six Self-Concept Measures

Split-half reliability coefficients were obtained for all six measures, and test-retest reliability coefficients were obtained for the three discrepancy measures. The period intervening between the test and the retest varied from *S* to *S*. The minimum was 228 days; the maximum was 420 days; the mean was 330 days. Using the Spearman-Brown correction, the split-half reliability coefficients for the three discrepancy variables were: "Self"–"Ideal Self" = + .88; "Social Self"–"Ideal Self" = + .85; "Self–"Social Self" = + .91; those for the three instability vari-

---

[3] Six of the nine correlations between discrepancy and instability measures have a minor degree of nonindependence. For example, one of the two sets of ratings contributing to an *S*'s score on discrepancy between "Self" and "Ideal Self" is the same as one of the three sets of ratings contributing to his "instability of Self" score, viz., his ratings of "Self" under placebo. However, it is not clear how this type of nonindependence could spuriously inflate the correlations involved. Furthermore, of the three correlations which *are* completely independent, one reaches the .01 level of significance and one reaches the .05 level of significance, using a two-tailed test.

Table 2

Correlations Between Six Self-Concept Variables and 49 Non-Self-Concept Variables

Measure	Discrepancy Variable			Instability Variable		
	Self–Ideal	Ideal–Social	Self–Social	Inst. Self	Inst. Social	Inst. Ideal
Self-insight measure	−.57**	−.41*	−.58**	−.55**	−.70**	−.43*
Average-mood measures						
Sleepy	+.06	−.10	+.21	+.30	+.28	+.03
Uneasy	+.37	+.17	+.46*	+.53**	+.59**	+.46*
Mind slow	+.16	+.13	+.24	+.48*	+.32	+.29
Dreamy	+.18	−.07	+.23	+.28	+.36	+.21
Sad	+.26	+.07	+.29	+.36	+.41*	+.47*
Grouchy	+.41*	+.22	+.42*	+.39	+.60**	+.51*
Groggy	+.30	+.14	+.32	+.48*	+.49*	+.27
Shy	+.31	+.11	+.28	+.32	+.48*	+.40*
No pep	+.35	+.20	+.34	+.46*	+.30	+.22
Worried	+.37	+.09	+.52**	+.38	+.50*	+.30
Restless	+.39	+.20	+.48*	+.35	+.39	+.30
Unsociable	+.15	−.03	+.21	+.19	+.30	+.33
Manson maladjustment	+.67**	+.37	+.64**	+.59**	+.40*	+.46*
Maslow insecurity	+.72**	+.44*	+.65**	+.55**	+.49*	+.43*
Edwards variables						
Achievement	−.10	−.19	−.26	−.24	−.13	−.39
Deference	−.31	−.03	−.28	−.02	+.05	+.22
Order	−.02	+.07	−.07	−.28	−.16	+.03
Exhibition	−.05	−.08	+.06	+.14	.00	−.23
Autonomy	+.05	+.17	−.03	+.13	+.06	+.14
Affiliation	−.02	−.01	+.09	−.11	−.22	−.05
Intraception	−.28	−.12	−.15	+.09	−.23	+.43*
Succorance	+.45*	+.41*	+.47*	+.22	+.16	+.17
Dominance	−.18	−.39	−.15	−.10	+.06	−.45*
Abasement	+.29	+.21	+.24	+.23	+.34	+.21
Nurturance	+.19	+.26	+.18	−.03	+.02	+.23
Change	−.38	−.34	−.35	+.11	−.14	+.13
Endurance	−.24	−.37	−.32	−.40*	−.19	−.36
Heterosexuality	+.18	+.29	+.18	+.17	+.12	−.01
Aggression	+.21	+.04	+.23	+.34	+.32	−.10
TSE extraversion						
Thinking	−.12	+.05	−.01	+.16	.00	+.17
Social	−.52**	−.66**	−.30	−.41*	−.34	−.41*
Emotional	+.27	+.18	+.22	−.01	+.08	+.26
Cattell factors						
A—Cyclothymia	−.27	−.29	−.03	−.08	−.11	−.08
B—Intelligence	+.63**	+.49*	+.43	+.24	+.38	+.05
C—Emot. stability	−.55*	−.44*	−.33	−.10	−.12	+.14
E—Dominance	−.37	−.39	−.20	−.17	−.54*	−.32
F—Surgency	−.22	−.31	−.06	−.13	+.08	−.03
G—Charac. strength	+.03	+.05	−.14	−.12	−.14	+.04
H—Adventurous	−.38	−.36	−.14	−.15	−.20	−.04
I—Emot. sensitivity	+.47*	+.21	+.37	+.25	+.13	+.02
L—Paran. schizothym.	+.19	+.25	+.23	+.40	+.05	+.26
M—Bohemianism	+.44*	+.28	+.46*	+.24	+.12	+.14
N—Sophistication	+.15	+.30	+.24	+.35	−.08	+.24
O—Anxious insecurity	+.60**	+.42	+.41	+.28	+.26	+.34
Q_1—Radicalism	+.29	+.16	+.16	−.16	−.02	−.21
Q_2—Independent	+.45*	+.61**	+.29	+.27	+.09	+.41
Q_3—Will control	−.42	−.38	−.38	−.46*	−.37	−.39
Q_4—Nervous tension	+.48*	+.48*	+.38	+.34	+.13	+.46*

Note.—For correlations involving Cattell Factors, $N = 20$. For all other correlations, $N = 24$.
* Significant at .05 level by two-tailed test. ** Significant at .01 level by two-tailed test.

ables were: "Self" = + .78; "Social Self" = + .86; "Ideal Self" = + .75. The test-retest coefficients for the three discrepancy variables were: "Self"–"Ideal Self" = + .82; "Social Self"–"Ideal Self" = + .63; "Self"–"Social Self" = + .78. All of these correlations were plotted to be certain that they were not being spuriously increased by a few *S*s with extreme scores. It was found that the correlations held throughout the scattergrams.

In a subsequent study of the reliability of the three discrepancy variables, in a group of *S*s whose discrepancy scores covered narrower ranges[4] than those covered by the 24 *S*s in the present study, somewhat lower reliability coefficients were obtained. In the subsequent experiment, the three discrepancy variables yielded split-half coefficients of + .79, + .83, and + .75. With an interval of two weeks, the test-retest coefficients were + .77, + .77, and + .60. The self-concept test is now undergoing item analysis in an effort to increase its reliability when used with *S*s whose scores cover only moderate ranges.

Correlations Between the Six Self-Concept Measures and Other Personality Measures

Table 2 reports correlations between each of the six self-concept variables and 49 "non-self-concept" measures of personality. A striking characteristic of the correlations reported in Table 2 is the consistency of the association between measures of adjustment and measures of self-concept discrepancy and instability. These correlations indicate that high instability scores and high discrepancy scores are associated with poor adjustment scores, whereas low instability scores and low discrepancy scores are associated with good adjustment scores.

The consistent correlation between measures of self-concept discrepancy and instability on the one hand, and measures of adjustment on the other hand, suggests one reason for the intercorrelation *among* the self-concept variables—namely, that the six self-concept variables are correlated with each other because all of them are correlated with a common variable: adjustment.

Discussion

Psychological Meaning of the Six Self-Concept Measures

Since the meaning of the correlations reported in Tables 1 and 2 depends on the meaning of the six self-concept variables involved in these correlations, it seems desirable to specify what we at present assume to be the psychological meaning of these six variables. The meaning of each variable will be given by indicating the psychological meaning which we attribute to a high score on that variable.

1. A high discrepancy between "Self" and "Ideal Self" indicates that the *S* feels inadequate relative to his ideal; that is, he evaluates himself unfavorably.
2. A high discrepancy between "Social Self" and "Ideal Self" indicates that the *S* feels that other people perceive him in ways which are very different from *his* standard of perfection; that is, he feels that their perceptions of him are, relative to his own ideal, unfavorable.
3. A high discrepancy between "Self" and "Social Self" indicates that the *S* thinks that other people do not accurately perceive and understand him.
4. High "instability of Self" indicates that the *S*'s attitudes toward himself undergo marked change from time to time.
5. High "instability of Social Self" indicates that the individual changes markedly from time to time concerning what he believes other people think of him.
6. High "instability of Ideal Self" indicates that the individual's goals, values, and ideals fluctuate markedly over time.

Interrelations Among the Six Self-Concept Measures

If the above attribution of meaning to the six self-concept variables is accepted as reasonable, then the positive correlations among these six variables become quite reasonable.

---

[4] For example, in the subsequent experiment, the *S* with the highest total discrepancy between "Self" and "Ideal Self" had an average item discrepancy of 1.36 scale points, and the *S* with the lowest total discrepancy between "Self" and "Ideal Self" had an average item discrepancy of 0.32 scale points. In contrast, in the experiment reported in the present paper, the highest average discrepancy between "Self" and "Ideal Self" was 1.93 scale points, and the lowest was 0.14 scale points.

It is not surprising, for example, that the *S* who evaluates himself unfavorably (high discrepancy between "Self" and "Ideal Self") also tends to react in the following ways: he feels that he is evaluated unfavorably by others (high discrepancy between "Social Self" and "Ideal Self"); he feels misperceived by others (high discrepancy between "Self" and "Social Self"); he changes markedly from time to time concerning what he thinks he is (high "instability of Self"); he changes markedly from time to time concerning what he thinks other people think of him (high "instability of Social Self"); and he changes markedly from time to time concerning what he would like to be (high "instability of Ideal Self"). Similarly, it is not surprising that the *S* who evaluates himself favorably also feels evaluated favorably by others, also feels understood by others, and also has rather stable ideas about what he is, what he believes others think of him, and what he would like to be.

Stating that the intercorrelation among the six self-concept variables is reasonable and is not due to experimental artifacts does not *explain* this intercorrelation. To attempt a theoretical explanation[5] of the intercorrelation among the six self-concept variables would involve a rather extensive manipulation of some empirical findings and many hunches. These findings and hunches would perhaps most profitably deal with the events occurring early in childhood which might affect the development of the total configuration of ideas and valuations which the individual has concerning himself. An extensive endeavor of this sort does not seem appropriate in the present report.

Dynamics. It might be useful, however, to mention just a few observations and hunches (suggested by numerous observers of human behavior: W. James, G. H. Mead, P. Lecky, C. Rogers, H. S. Sullivan, K. Horney, to mention a few) which might be helpful in developing such an explanation. For example, the statements presented in the following three paragraphs might, if they are valid, shed some light on the factors involved in the production of the correlations between the "Social Self"–"Ideal Self" discrepancy and the other five measures of self-concept discrepancy and instability. (Earlier in this section we assumed that a high discrepancy between "Social Self" and "Ideal Self" indicates that the *S* feels that the perceptions of him held by other people are, relative to *his own* standard of perfection, unfavorable. For the sake of speculation, we now assume that the degree of discrepancy between "Social Self" and "Ideal Self" reflects the degree to which the individual *feels valued* by other people.) The following statements are presented merely to exemplify the way in which speculation about events which might influence the development of various aspects of the self-concept might give some meaning to the five correlations involving the "Social Self"–"Ideal Self" discrepancy. (The remaining 10 correlations will not be treated in this manner in the present report.)

Concerning the correlation between the "Social Self"–"Ideal Self" discrepancy and the "Self"–"Ideal Self" discrepancy, where $r = +.74$, one might speculate as follows: The child's ideas about, and valuations of, himself are profoundly influenced by what he believes to be the ideas held about him and the valuations made of him by other people with whom he interacts. That is, early in life the individual learns what to think of himself by observing what others think of him. If one feels highly valued by others, he is likely to value himself highly. If one does not feel highly valued by others, he is not likely to value himself highly.

Concerning the correlation between the "Social Self"–"Ideal Self" discrepancy and the "Self"–"Social Self" discrepancy, where $r = +.51$, one might say: To feel valued by others is a powerful human motive. If an individual feels that he is not valued by others, he may attempt to deal with the disappointment by concluding that he is not valued because he is not understood.

Concerning the correlations between the "Social Self"–"Ideal Self" discrepancy and the instability of "Ideal Self," "Self," and "Social Self," where the r values are + .31, + .45, and + .47, respectively, one might say: The frustration of an individual's desire to feel valued by others, and the consequent attempts to understand *why* he is not valued, could lead to uncertainty and confusion concerning (*a*) the ideals and standards he should try to maintain, (*b*) the qualities and characteristics which he feels comprise his "Self," and (*c*) the validity of his notions of what other people think of him.

---

[5] Following Feigl's (11) definition of explanation as being the "inductive-deductive or (on higher levels) hypothetico-deductive derivation of the more specific (ultimately descriptive) propositions from more general assumptions (laws, hypotheses, theoretical postulates) in conjunction with other descriptive propositions (and often together with definitions)."

Additional data needed. Before engaging in further speculation concerning the possible meanings of the positive correlations among the six self-concept variables, it seems desirable to obtain considerably more empirical data dealing with such questions as: Are the correlations reported in the present paper repeatable; and, if so, in what population of *S*s? What personality characteristics distinguish individuals who contribute to these positive correlations from individuals who do not? In what ways are the six self-concept variables similar to and different from each other in meaning, i.e., what personality variables correlate similarly with the six self-concept variables with respect to magnitude and direction; and what personality variables correlate differentially with the six self-concept variables with respect to magnitude and direction? How do these six self-concept variables relate to other self-concept dimensions? It might be particularly worthwhile to determine the way in which these six variables relate to a measure of Lecky's (14) concept of "self-consistency." What are the developmental relations among the six self-concept variables? For example, one might expect the development of the ideas and attitudes reflected by the "Social Self"–"Ideal Self" discrepancy to be important in affecting the development of those reflected by the other five variables. Do the instability variables reflect the degree of unclarity of ideas and attitudes relating to the self; or do they reflect conflicting and mutually inconsistent ideas and attitudes; or do they reflect both; or do they reflect neither; or do they reflect one in some *S*s, the other in some *S*s, both in some *S*s, and neither in some *S*s?

Relations Between the Six Self-Concept Variables and Other Personality Measures.

At present, perhaps the most useful way of giving meaning to the six self-concept variables is by specifying the way in which they are correlated with other measures of personality which have already acquired meaning. Table 2 presents these correlations. Ideally, the meaning of the "non-self-concept" measures should be given by specifying the items on which they are based and by presenting the validity information concerning these measures. However, space limitations prevent this in the present paper.

In the results section, it was stated that there is a highly consistent tendency for high discrepancy and high instability of self-concept to be associated with poor adjustment and for low discrepancy and low instability to be associated with good adjustment. The acceptability of this statement depends on the acceptability of the assumption that the "non-self-concept" measures, which are correlated with the self-concept variables, do indeed measure some aspect of maladjustment-adjustment. In the present paper, extensive effort to justify this assumption will not be made.

Self insight. Table 2 indicates that all six self-concept measures are negatively correlated with self-insight at or beyond the .05 level of confidence, using a two-tailed test. That is, high discrepancy on all three discrepancy measures and high instability on all three instability measures are associated with low self-insight. This finding is congruous with Rogers' conception (20) of the relations among adjustment, self-awareness, and discrepancy between "Self" and "Ideal Self."

Average mood. Table 2 shows that 19 of the 72 correlations between self-concept measures and measures of average-mood are significant at or beyond the .05 level, using a two-tailed test. These 19 correlations indicate (and the remaining 53 suggest) that high self-concept discrepancy and high self-concept instability are associated with reporting feeling uneasy, grouchy, worried, nervous, groggy, shy, sad, not peppy, mentally slow, dreamy, unsociable, and sleepy. Of the 53 correlations between self-concept variables and average-mood variables which do not reach the .05

level of significance, only three fail to be consistent with the statement that high self-concept discrepancy and instability are associated with the mood characteristics specified in the previous sentence. These 72 correlations yield an unusually consistent picture. If an *S*'s ratings of his mood on the 12 mood items on the 15 different occasions are taken as valid indications of the way he generally feels, the correlations between self-concept measures and average-mood measures indicate that *S*s with high self-concept discrepancy and high self-concept instability feel more anxious, irritable, unhappy, and inadequate than do *S*s with low self-concept discrepancy and low self-concept instability.

These correlations between self-concept scores and average-mood scores are quite reasonable. It is not surprising that the person who disapproves of himself (high discrepancy between "Self" and "Ideal Self") *and* feels others disapprove of him (high discrepancy between "Social Self" and "Ideal Self") *and* feels misunderstood by others (high discrepancy between "Self" and "Social Self") *and* is unable to maintain a stable notion of what he is, what he wants to be, or what he thinks others think of him (high instability of "Self," "Ideal Self," and "Social Self") would feel anxious, irritable, unhappy, and inadequate. Similarly, it is not surprising that the person who approves of himself *and* thinks others approve of him *and* feels understood by others *and* has a stable notion of what he is, what he wants to be, and what he thinks others think of him, would feel secure, friendly, happy, and adequate.

Manson and Maslow questionnaires. The correlations between self-concept measures and the Manson and Maslow tests indicate that high self-concept discrepancy and instability are associated with high "Manson insecurity" and high "Maslow maladjustment." Eleven of these twelve correlations are significant at or beyond the .05 level of confidence, using a two-tailed test. The *z* transformed average *r* for these 12 correlations is + .55. Thus, it is clear that the aspects of adjustment-maladjustment which are measured by the Manson and Maslow tests are very similar to those aspects of adjustment-maladjustment which are measured by the six self-concept variables.

Edwards variables. Most of the correlations between the self-concept variables and the Edwards variables are not high enough to be significant at the .05 level using a two-tailed test. However, when the total configuration of correlations between the Edwards variables and the self-concept variables is examined, it is seen that high discrepancy and high instability consistently tend to be associated with need characteristics which are frequently found in poorly adjusted individuals. (The correlation between "instability of Ideal Self" and need intraception may be an exception.) Need succorance, need dominance, need endurance, and need intraception are associated with one or more of the six self-concept variables at the .05 level of confidence using a two-tailed test. High instability and discrepancy are associated with high need succorance, low need dominance, low need endurance, and high need intraception. Although not significant at the .05 level of confidence, the correlations between the self-concept scores and scores on need abasement, need achievement, and need aggression indicate that high instability and discrepancy might be associated with high need abasement, low need achievement, and high need aggression.

T-S-E measures. Only the Social scale in the Minnesota Thinking, Social, and Emotional Introversion-Extraversion Test correlates significantly with the self-concept variables. Four of the six self-concept variables correlate with the Social Introversion-Extraversion variable at or beyond the .05 level of confidence, using a two-tailed test. High self-concept discrepancy and instability are associated with social introversion.

Cattell factors. The correlations between the Cattell Factors and the self-concept variables agree with the findings discussed above in that high self-concept discrepancy and instability are associated with poor adjustment. Table 2 indicates that one or more of the six self-concept variables is significantly positively correlated with Factors M, O, and Q_4, and significantly negatively correlated with Factors, C, E, and Q_3. Although not significant at the .05 level of confidence, there is a consistent tendency for the self-concept variables to be positively correlated with Factor L. In other words, high discrepancy and instability are associated with high Factor M

("Hysteric Unconcern" or "Bohemianism"), high Factor O ("Anxious insecurity"), high Factor Q_4 ("Nervous tension"), low Factor C ("Emotional stability"), low Factor E ("Dominance or ascendance"), low Factor Q_3 ("Will control and character stability"), and high Factor L ("Paranoid schizothymia"). Cattell has recently reported (6) that five of these factors (viz., O, Q_4, C, Q_3, and L) are found to comprise a second-order factor, which he names Anxiety vs. Dynamic Integration. Cattell suggests three alternative interpretations of this second-order factor: "first, that it is a factor of general anxiety; second, a factor of general neuroticism; and third, that it is the different form of personality disorganization associated with psychoticism" (6, p. 416). In this second-order factor, the primary factors O, Q_4, and L have positive signs, while Q_3 and C have negative signs. Of the 30 correlations between the six self-concept variables and the five primary factors comprising this second-order factor, 29 are consistent with the conclusion that high scores on the six self-concept variables are indicative of poor adjustment. Seven of these 29 correlations are significant at or beyond the .05 level using a two-tailed test.

General comments. The principal findings of the present study are that the six self-concept variables are positively correlated with each other and are negatively correlated with adjustment. The high intercorrelation among the six self-concept variables (Table 1) and the consistency of their correlation with various measures of adjustment (Table 2) help clarify the meaning of these six self-concept variables by indicating the ways in which they are similar. However, clarification of meaning can also be obtained by indicating the dissimilarities among these six variables. Data obtained from a sample considerably larger than that of the present study, analyzed with factorial techniques, would be useful in this respect.

Even within the present data, however, there are some suggestions concerning the ways in which the six self-concept variables might differ. For example, Table 1 indicates that "instability of Ideal Self" is probably the most unique of the six variables. Table 2 also offers some suggestions concerning possible differences among these six variables. For example, the Cattell factors correlate higher with the discrepancy variables than with the instability variables. This is particularly true of the correlations involving the discrepancy between "Self" and "Ideal Self." On the other hand, the average-mood variables correlate higher with the instability than with the discrepancy variables. This is particularly true of the correlations involving the "instability of Social Self."

Theorists and experimenters have found the discrepancy between "Self" and "Ideal Self" to be a useful measure. Perhaps further study of the other five self-concept measures discussed in the present paper would show them also to be of value. This would be particularly likely in the case of the discrepancy between "Social Self" and "Ideal Self," if it were found that the development of the ideas and attitudes reflected by this measure is important in affecting the development of the ideas and attitudes reflected by the other five self-concept variables.

Further study of the three instability variables also seems worthwhile. Inability to maintain a stable conception of the "Self," "Ideal Self," and "Social Self" is associated with low self-insight and with other aspects of maladjustment, such as anxiety, hostility, dejection, and feelings of inadequacy. A thorough understanding of this association should be useful in the further development of personality theory and therapeutic technique. If it were found, for example, that individuals with high "instability of Self" scores allow their self-evaluations to be inordinately influenced by immediately preceding successes, failures, approvals, disapprovals, etc., whereas individuals with low scores anchor their self-evaluations firmly in the events of their total past, then the meaning of the association between "instability of Self" and various measures of adjustment would become clarified.

Summary

Scores on six self-concept measures were obtained for each of 24 college men. These six measures were: discrepancy between "Self" and "Ideal Self," discrepancy between "Self" and "Social Self," discrepancy between "Social Self" and "Ideal Self," "instability of Self," "instability of Ideal Self," and "instability of Social Self." All six measures were

found to be positively correlated with each other. Eleven of the 15 correlations were significant at the .05 level of confidence using a two-tailed test. The possibility of spuriously high intercorrelation among the six self-concept measures was examined, and the reasons for rejecting this possibility were discussed.

Split-half reliability coefficients were obtained and reported for all six measures, and test-retest coefficients were obtained and reported for the three discrepancy measures.

The psychological meanings presently attributed to the six self-concept measures were specified, and it was concluded that in view of these meanings, the intercorrelation among the six measures is reasonable. Five of the 15 correlations among the six measures (namely, those involving the "Social Self"–"Ideal Self" discrepancy) were discussed in some detail, and dynamics which could give rise to these five correlations were suggested. Certain questions which appear to be worthy of experimental examination were raised concerning the meaning of the six measures and concerning the meaning of their interrelations.

The correlations between each of the six self-concept measures and each of 49 other measures of personality were presented and discussed. One of these 49 was a measure of self-insight; 12 were measures of average-mood; the remaining 36 measures were obtained with personality inventories. On the basis of these 294 correlations, it was concluded that all six self-concept variables are highly related to such aspects of adjustment as self-insight, anxiety, dejection, friendliness, insecurity, social introversion, will control, etc. High discrepancy scores on all three discrepancy measures, and high instability scores on all three instability measures were found to be associated with poor adjustment scores.

It was suggested that one possible reason for the intercorrelation among the six self-concept variables is their correlation with a common variable: adjustment.

Received May 3, 1957.

References

1. Bills, R. E., Vance, E. L., & McLean, O. S. An index of adjustment and values. *J. consult. Psychol.*, 1951, **15**, 257–261.
2. Berger, E. N. The relationship between expressed acceptance of self and expressed acceptance of others. *Univer. Pittsburgh Bull.*, April 10, 1951, **47**, 1–8.
3. Brownfain, J. J. Stability of the self-concept as a dimension of personality. *J. abnorm. soc. Psychol.*, 1952, **47**, 597–606.
4. Cattell, R. B., Saunders, D. R., & Stice, G. *The 16 P.F. Test*. Champaign, Ill.: Institute for Personality and Ability Testing, 1950.
5. Cattell, R. B. Confirmation and clarification of primary personality factors. *Psychometrika,* 1947, **12**, 197–220.
6. Cattell, R. B. Second-order personality factors in the questionnaire realm. *J. consult. Psychol.*, 1956, **20**, 411–418.
7. Cowen, E. L. The negative self-concept as a personality measure. *J. consult. Psychol.*, 1954, **18**, 138–142.
8. Edwards, A. L. *Edwards Personal Preference Schedule*. New York: Psychological Corp., 1953.
9. Edwards, A. L. *Edwards Personal Preference Schedule* (Manual). New York: Psychological Corp., 1954.
10. Evans, C., & McConnell, T. R. *Minnesota T-S-E Inventory*. Chicago: Science Research Associates, 1942.
11. Feigl, H. Some Remarks on the Meaning of Scientific Explanation. In H. Feigl & W. Sellars (Eds.), *Readings in Philosophical Analysis*. New York: Appleton, Century, Crofts, 1949.
12. Gravenstein, J. S., Smith, G. M., Sphire, R. D., Isaacs, J. P., & Beecher, H. K. Dihydrocodeine: Further development in measurement of analgesic power and appraisal of psychologic side effects of analgesic agents. *New England J. Med.*, 1956, **254**, 877–885.
13. Lasagna, L., vonFelsinger, J. M., & Beecher, H. K. Drug-Induced Mood Changes in Man. I. Observations on healthy subjects, chronically ill patients and post-addicts. *J. Amer. med. Ass.*, 1955, **157**, 1006–1020.
14. Lecky, P. *Self-Consistency: A theory of personality*. New York: Island Press, 1945.
15. Manson, M. P. *The Manson evaluation*. Beverly Hills, Calif.: Western Psychol. Services, 1948.
16. Maslow, A. H. *The S-I inventory*. Stanford University, Calif.: Stanford University Press, 1945.
17. Murray, H. A. *Explorations in personality*. New York: Oxford Univer. Press, 1938.
18. Phillips, E. L. Attitudes toward self and others: A brief questionnaire report. *J. consult. Psychol.*, 1951, **15**, 79–81.
19. Rogers, C. R. *Client-centered therapy*. Boston: Houghton Mifflin, 1951.
20. Rogers, C. R., & Dymond, R. F. *Psychotherapy and personality change*. Chicago: Univer. of Chicago Press, 1954.
21. Smith, G. M., & Beecher, H. K. Quantitative determination of the effects of drugs on mood and sensation. Unpublished manuscript, Anesthesia Department, Massachusetts General Hospital, Boston.

Journal of Consulting Psychology
Vol. 22, No. 2, 1958

Differentiation of Clinical Groups Using Canonical Variates[1]

H. R. Beech and A. E. Maxwell
Institute of Psychiatry, University of London

The present study is part of an investigation into the factors which influence the reproduction of simple designs by drawing, and represents an extension of work by Shapiro (11, 12, 13) and Yates (14) into rotation effects. The generalizations put forward by Shapiro to account for the greater amounts of rotation found among brain-damaged subjects (13) were used in predicting certain other anomalies which would be found in the drawing reproductions of that group. The main notion which led to the measures (i.e. measures other than the rotation of designs) being taken in the recent study was that exaggerated negative induction effects result in the restriction of the available perceptual cues in brain-damage patients.

This paper presents the results obtained from a canonical variate analysis employing four measures or tests derived from the application of Shapiro's generalizations about the effects of brain damage to the Drawing Rotation Test (14). The main aim of the analysis was to improve the differentiations obtained between brain-damaged and nonorganic groups. Other studies (4, 5) making use of this statistical technique have produced favourable results so far as differentiation is concerned.

The Measures

The Drawing Rotation Test comprises 40 model designs which are placed in front of the subject (*S*) at specified distances, and he is asked to copy them by drawing. All the designs are simple, symmetrical, and capable of reproduction using four Kohs blocks. Ten different designs are used in all, each of which appears four times in the test series, once in each of the four possible combinations of figure and ground orientations. Three examples of these designs are shown in Fig. 1. The models may be in either square or diamond orientation, and it has been found (12, 13) that organic *S*s show a tendency to rotate their reproductions (i.e. to produce discrepancies in orientation between the model and their reproduction), which is greater than that found among nonorganic patients. Our *first* measure, therefore, was an index of the extent to which reproduction misorientation occurred, this index being the mean rotation score for the 40 designs given. We have called this measure Test 3. A measure of the time taken to complete the first eight designs was called Test 4. In the Rotation Test an *S* is allowed up to three minutes in which to complete any design, although failure to do so in the given time is very rare. The time measure, which has been found to differentiate between normals and abnormals (2), was the mean time, in seconds, which the *S* took to complete the first eight designs out of the series of forty. The decision to consider only eight designs for Test 4 was quite arbitrary and purely a matter of convenience in mak-

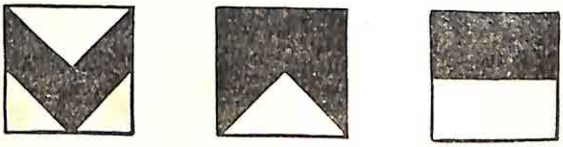


Fig. 1. Three examples of the designs used in the study.

---

[1] The work reported in this paper was made possible by a grant from the Research Fund, made available from the endowment by the Board of Governors of the Bethlem Royal and Maudsley Hospitals.

ing the necessary measurements. This limitation of the number of designs also applied to Tests 1 and 2 where the scoring procedure was too time-consuming to permit consideration of a large number of reproductions.

Test 1, called disproportionality (2), may be looked upon as an index of the magnitude of errors in bisection. Almost all the designs used in the test have their internal structure so arranged that certain lines exactly bisect the external framework of the design. Disproportionality represents a ratio of parts of the design which, because of bisecting internal parts, should be equal in length. Lack of proportionality, or disproportionality, was measured by dividing the longer of the two lengths which should be equal by the shorter. In all, twenty-five of these indices were obtained from the first eight designs, the mean disproportionality score being the sum of the ratios (taken to three decimal places) divided by 25.

Test 2, colour score, represents an attempt to estimate the extent to which the yellow aspects of an *S*'s reproductions occupy a larger amount of the available space than the blue aspects. All the designs are in blue and yellow. The *S* is asked not only to draw the design but also to indicate which parts are blue and which yellow in his reproduction. The measurements taken are based upon the fact that in each design equal amounts of space are given to the two colours, so that any tendency to exaggerate yellow aspects at the expense of blue will be discovered by taking a ratio of blue to yellow parts. In practice, this effect is tapped by comparing blue and yellow lengths where these colours form borders to the external framework of the design, always dividing the yellow length by the blue.

Both colour and disproportionality scores have similarities in that both deal with parts of the designs which should be equal and both involve ratios of these parts, but it has been shown (2) that there is no necessary correspondence between the two measures.

The Samples

Five groups of *S*s were used in the statistical analysis of the four measures, 48 normals, 20 neurotics, 26 schizophrenics, 17 depressives, and 33 brain damaged, making a total of 144 *S*s. Most of these *S*s were tested by three other investigators in the Maudsley Hospital as part of their research program. We are indebted to these experimenters for allowing us to make use of the drawings collected by them. Special care was taken in the case of functional patients and normal *S*s to exclude those persons with symptoms or history which were consistent with the possibility of brain damage.

The normals comprised *S*s drawn from adult education classes, nurses, secretarial staff, and other miscellaneous occupations.

The brain-damaged group could be divided into two subgroups: a group of patients from the Neurosurgical Unit of the Maudsley Hospital whose brain injury was a result of surgical interference (e.g., temporal lobectomy cases), and a group whose brain injury was due to some disease process (e.g., G.P.I.). It was found that these two categories did not differ significantly in terms of their drawing test scores.

The depressives were not all psychotic conditions but included some neurotic depression cases. The main distinguishing feature of this group was that all the individuals comprising the group were to receive ECT as part of their treatment to relieve a depressed condition.

The schizophrenics included early and chronic cases in about equal proportions. They were drawn mainly from the Maudsley Hospital.

The neurotic group was composed almost entirely of patients diagnosed as "hysteric" or "anxiety state." All of these *S*s were drawn from the patient population of the Bethlem and Maudsley Hospitals.

In all, five groups of *S*s were tested by four different experimenters at more than one hospital.

The Problem

Comparing the five groups on the four Drawing Test measures produced significant differences between them. The statistical details of these differences are shown in Table 1. It will be seen that where disproportionality (Test 1), colour (Test 2), and rotation (Test 3) are concerned the differences are mainly between brain-damaged and non-brain-damaged groups, while on time (Test 4) the differentiation is mainly between normal and abnormal groups.

However, although significant differences between the groups on these measures are obtained, we are left with large areas of misclassification, and the results have limited

Table 1

Significance of Differences Between the Groups on the Drawing Test Measures

F ratios			
Test 1	Test 2	Test 3	Test 4
9.379**	8.687**	3.743**	19.613**

	t (2-tail)			
Groups compared	Test 1	Test 2	Test 3	Test 4
Brain damaged v. normal	3.538**	4.192**	3.337**	7.920**
Brain damaged v. depressive	2.547*	2.488*	1.714	1.731
Brain damaged v. schizophrenic	4.726**	5.446**	2.378*	1.150
Brain damaged v. neurotic	3.856**	4.006**	3.217**	3.252**
Normal v. depressive	0.139	0.757	0.977	4.601**
Normal v. schizophrenic	1.773	1.891	0.541	6.220**
Normal v. neurotic	1.088	0.662	0.587	3.357**
Depressive v. schizophrenic	1.514	2.176*	0.458	0.689
Depressive v. neurotic	0.477	1.187	1.310	1.225
Schizophrenic v. neurotic	0.477	0.477	0.968	2.081*

Note.—Tests 1, 2, 3, and 4 refer to the measures Disproportionality, Colour, Rotation, and Time, respectively.
* Significance of .05 or above.
** Significance of .01 or above.

clinical value. Even in the case of brain-damaged subjects, the outcome is disappointing, for the bulk of this group overlaps with other groups.

The main task at this stage, therefore, was to attempt to improve the differentiation between the groups. The method chosen to achieve this end was, as stated above, a discriminant function analysis. The statistical model employed in such an analysis is now briefly described.

The Statistical Model

The model is one of finding a linear function, or orthogonal linear functions, of the test scores which will best discriminate between the groups. If x_1 to x_4 are taken to represent the four tests and u_1 to u_4 are weights, such a function has the form:

$$f(x) = u_1x_1 + u_2x_2 + u_3x_3 + u_4x_4 \qquad (1)$$

It will be realised that when only two groups are concerned, a solution to the problem is provided by Fisher's linear discriminant function (6). When more than two groups are concerned, it cannot be assumed that the group means are colinear and a method is required which will indicate the number of dimensions in which the group means lie. One approach to the problem is that described by Rao (9) and used by Rao and Slater (10). The procedure involves maximising Mahalanobis' D^2. Another approach is that described by Lubin (7) and illustrated in a recent article by Eysenck (3). Here the square root of the correlation ratio, η, between the groups, on the one hand, and a linear function of the test scores, on the other, is maximised. This is the procedure followed here. It involves calculating a product-sum matrix, G, for the combined groups on the tests concerned, and a between-groups product-sum matrix, B. The problem then is to find a set of weights, i.e., a column vector, u, which will maximise the expression:

$$\eta^2 = u'Bu/u'Gu \qquad (2)$$

where u' is the transpose of u.

On differentiating this expression with respect to u and equating the result to zero, the determinental equation which gives the required values of u and the corresponding values of η^2 is obtained. It is:

$$(G^{-1}B - \eta^2 I)u = 0 \qquad (3)$$

where I is a unit diagonal matrix.

The solution of equation (*3*) involves finding the latent roots and latent vectors of the asymmetric matrix $G^{-1}B$ (8). The number of latent roots possible will depend on the rank of the matrix. This can be shown (7) to be the number of tests or one less than the number of groups, whichever is smaller.

It follows that when only two groups are being considered equation (*3*) will give, at most, one latent root. This is in agreement with the fact that one dimension is sufficient to account for two mean points. In our example, five groups and four tests are concerned, so that the rank of our $G^{-1}B$ matrix is four, and the matrix will have four latent roots. However, these roots may not all be significant. The number which is significant indicates the number of dimensions necessary and sufficient to account for the group means.

The assumptions involved when the above method is employed are not always clearly stated. In the first place, the variance–covariance matrices for the individual groups should be homogeneous, otherwise one is not justified in combining the groups when getting the total product-sum matrix. Secondly, it is required that the scores for the combined group on the several tests give a multivariate normal distribution; for unless this is so, one is not justified in setting up *linear* functions of the raw test.

Analysis of the Data

The original raw scores on each of the tests were not normally distributed and, after inspection, a logarithmic transformation was employed to stabilize their variances. This had the additional advantage of making the data more normal.

Inspection of the standard deviations, given in Table 2, reveals some discrepancies. In each test the standard deviation for the brain-damaged group is the largest and on the first test is roughly three times as large as that for any other group. To check on the significance of these differences, F ratios between variances on the tests were performed. These showed no differences between the variances on Test 4, time, but on Tests 1, disproportionality, 2, colour, and 3, rotation, the variances for the brain-damaged group were found to differ significantly from that of the group with the lowest variance on each of the tests. In other words, homogeneity of variance between the groups on Tests 1, 2, and 3 has not been achieved.

As a result, our data are not ideally suited to the statistical model proposed since the brain-damaged group has an advantage over the other two groups in that it is more highly weighted on account of its larger standard deviations. However, from a practical point of view, this may not be a bad thing since the brain-damaged patients are precisely those which we are most interested in differentiating from the others.

There is no simple way of testing for homogeneity of covariance between covariance matrices derived from small samples, but rank-order correlation coefficients between the tests for each sample separately and the product-moment correlations for the combined samples were calculated. The results gave no indication of gross departures from homogeneity of covariance between the groups. The product-moment correlations for the combined groups are given in Table 3.

In view of the findings in these preliminary investigations, it appeared reasonably legitimate, though not ideally so, to proceed with the calculations and obtain the latent roots and corresponding latent vectors of the ma-

Table 2

Means and Standard Deviations of Transformed Drawing Test Scores

Group	Test 1		Test 2		Test 3		Test 4	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Normal	0.1219	0.0584	1.0400	0.0548	0.6265	0.1882	0.0482	0.2004
Brain damaged	0.1870	0.1411	1.0948	0.0776	0.8297	0.3754	1.4399	0.2364
Depressive	0.1251	0.0520	1.0522	0.0400	0.7003	0.2817	1.3284	0.1995
Schizophrenic	0.0868	0.0420	1.0139	0.0362	0.6620	0.2442	1.3748	0.2166
Neurotic	0.0983	0.0384	1.0300	0.0436	0.5846	0.1894	1.2412	0.2004

Table 3

Correlations Between the Measures for the Total Group

Tests	Tests 1	2	3	4
1		0.62	0.59	0.30
2			0.44	0.12
3				0.30

trix $G^{-1}B$. The latent roots in order of extraction were 0.3849, 0.2231, and 0.0137; the percentage variance accounted for by each being 61.9, 35.9, and 2.2, respectively. The first two roots were found to be highly significant, while the third was not significant (1). This showed that the means of the groups on the tests could be represented adequately in a two-dimensional space.

The latent column vectors, corresponding to the first two latent roots, were then found. They are, respectively:

$$-0.2194 \qquad 1.0000 \qquad 0.0317 \qquad 0.6552$$

and

$$1.0000 \qquad 0.8018 \qquad -0.0746 \qquad -0.1922$$

Using these values as weights in the equation (*1*), two linear functions of the test scores are obtained:

$$X_1 = -0.2194x_1 + 1.0000x_2 + 0.0317x_3 + 0.6552x_4 \quad (4)$$

$$X_2 = 1.0000x_1 + 0.8018x_2 - 0.0746x_3 - 0.1922x_4 \quad (5)$$

On substituting, in turn, the scores of each individual in these equations, composite scores for each on the new variates X_1 and X_2 are obtained. These variates are sometimes called "canonical variates"; they are orthogonal to each other. The correlation ratio between the X_1 scores and the groups is given by the square root of the first latent root, namely, 0.6204. It is the maximum correlation ratio possible between the groups and any linear function of their scores. The X_2 scores, in turn, give a correlation ratio with the groups equal to the square root of the second latent root, that is, 0.4723.

In Fig. *2* the means of the canonical variate scores for the five groups are plotted. From this figure a picture of the relative positions of the group means is obtained.

Classification

At this stage, a further aspect of the problem of discriminating between the groups can be considered. This involves calculating for each individual the likelihood of his belonging to each of the five groups and assigning him to that group for which his likelihood is greatest. The statistical procedure for doing this is given by Rao (9), and in less advanced mathematical language by Lubin (7). By employing it, it is possible to partition either the original test space, or, alternatively, the canonical variate space between the groups. The calculations involved are heavy, but fortunately, as shown in Eysenck's paper (3), when only two variables are concerned the partitioning can be done satisfactorily by eye. Taking the original allocation to diagnostic categories as correct, lines radiating from a central point in the scattergram can be found which give a minimum of misclassification. For the sake of clarity, two scattergrams have been used here. Figure 3 contains the points for the normal, schizophrenic, and brain-damaged groups only, while Fig. 4 shows those for the depressive and brain-damaged groups only.

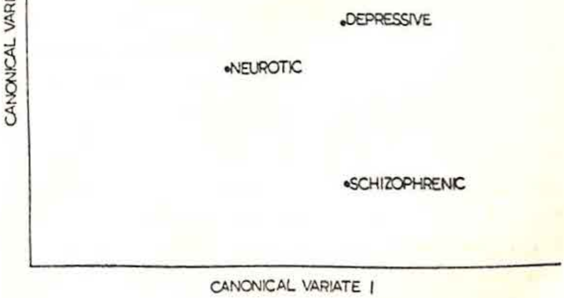


Fig. 2. Relative positions of the mean canonical variate scores for the five groups.

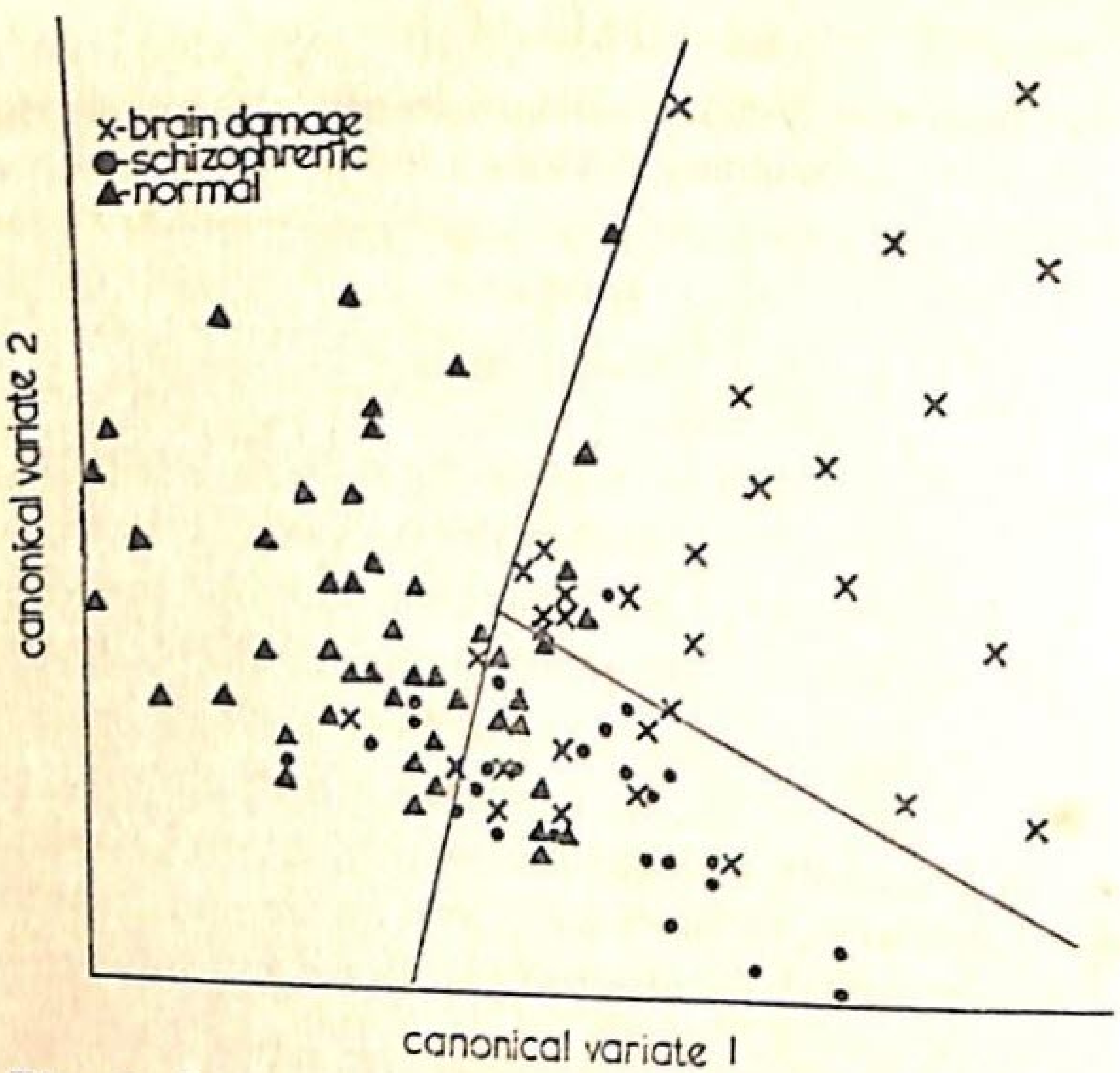


Fig. 3. Individual canonical variate scores for the normal, schizophrenic, and brain-damaged groups.

Discussion of the Results

The discriminant function analysis tells us the minimum number of independent dimensions necessary and sufficient to account for the information contained in the test results and assigns weights to the tests which maximize the differences between the group means along these dimensions. The test scores can now be replaced by the composite scores on the canonical variates—the means of which, for each group separately, are plotted in Fig. 2. The advantages of doing so are discussed below.

Comparing the *F* ratios and *t* tests for the individual measures (Table 1) with those for the canonical variates (Table 4), the superiority of the latter as an aid to differentiating between the groups may not immediately be obvious, for the levels of significance have not been increased very noticeably. However, a glance at Table 3 shows that the four tests are correlated positively and to a considerable extent with each other. As a result, the significance levels between the groups shown by the separate tests are not independent of each other and the amount of information available to us for discriminating between the groups appears greater than it really is. On the other hand, the two canonical variates are uncorrelated, so that the significance levels for the separate variates in Table 4 are independent. In other words, from the fact that Variate 1, for example, differentiates between two groups, we can draw no conclusion as to what result will be obtained when the same two groups are compared on Variate 2.

A further advantage of obtaining the canonical variates is that, whereas a four-dimensional space would be required to locate our *S*s if their scores on the four tests are used, the fact that only two latent roots were found to be significant means that the canonical variates allow us to locate the *S*s, without loss of information, in a two-dimensional space (see Fig. 2). The advantage thus achieved cannot be overemphasized.

Like factors in a factor analysis, canonical variates are, as seen from equation (*1*), weighted measures of the test scores. For this reason there is a natural tendency to try to interpret canonical variates psychologically, just as factors are interpreted. This tendency is not one to which the investigator should succumb without due caution. For one thing, invariance of factors from one study to the next is achieved by rotation, so that the factors rotated to a psychologically meaningful position are not *wholly* dependent upon the selection of tests used in the study. On the other hand, the weights used to define a canonical variate are unique to the study in question and to the tests and groups used in that study. Moreover, these weights, like regression weights in general, are known to

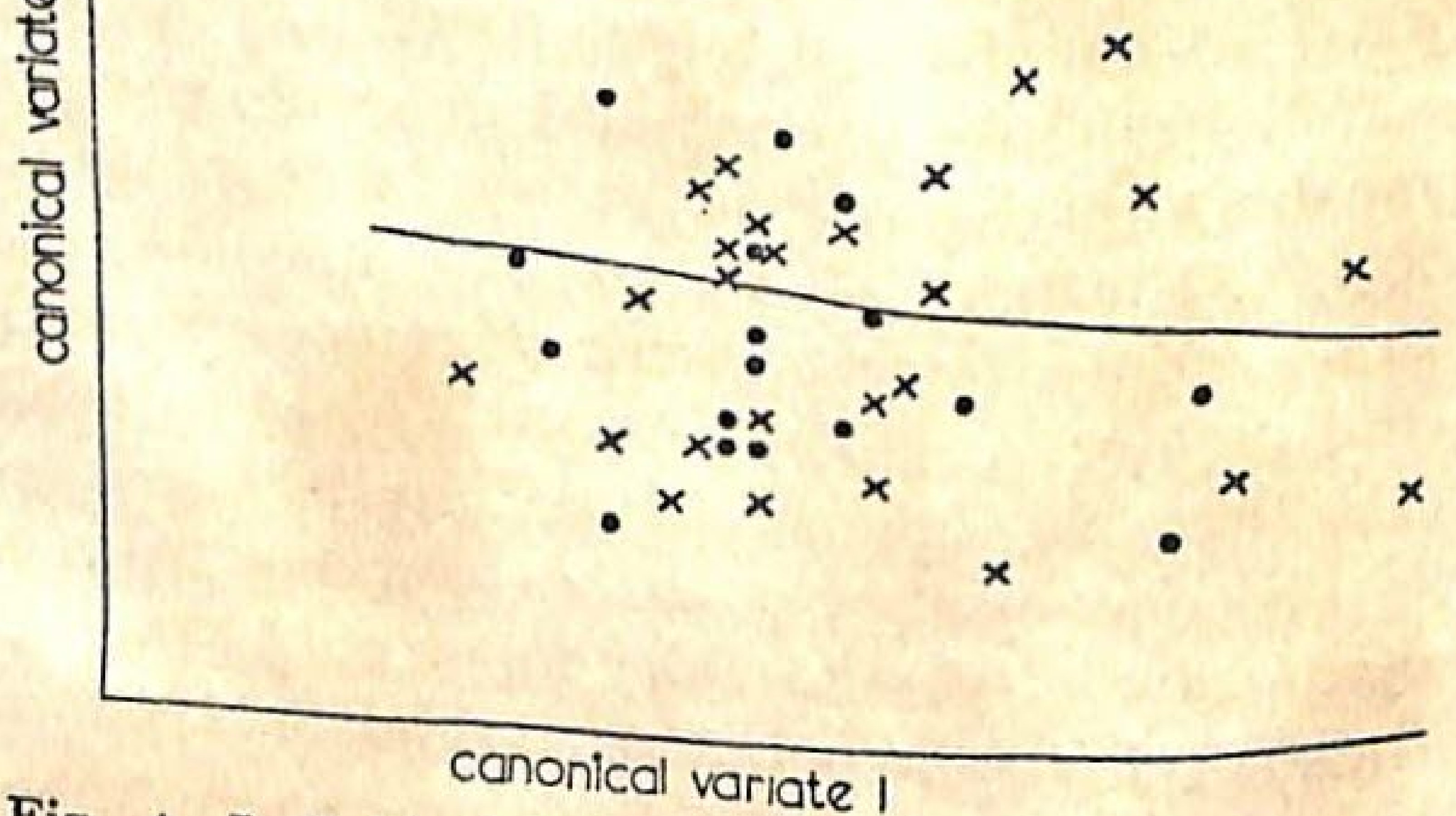


Fig. 4. Individual canonical variate scores for the depressive and brain-damaged groups.

vary notoriously from one study to another, especially so when the composition of the battery is altered. However, with this warning in mind, it will be helpful when discussing the differences between our groups to give psychological labels to the two canonical variates. For instance, canonical Variate 1, with high loadings on Tests 2 and 4, might tentatively be thought of as defining a normality-abnormality dimension, normals appearing at one extreme, neurotics in the middle, and the grossly abnormal groups at the other extreme. We have, however, no ready interpretation for the second canonical variate which is heavily weighted on Tests 1 and 2. Defining this dimension poses a problem toward which future research may be directed.

Now let us concentrate for a moment on Fig. 2. We may recall that the canonical variates, represented in it by the orthogonal axes, have the property of separating the group means from each other to the optimum degree. The mean points for the groups show that brain-damaged patients tend to have high scores on both canonical variates. Their mean is the most extreme of the means on both dimensions.

On the normality-abnormality dimension, the mean of the brain-damaged group occupies the most extreme position at the "abnormal" end, with the schizophrenic and depressive groups next, their means not differing significantly from that of the brain-damaged group. On the second dimension, the brain-damaged group mean together with the normal group mean are at one extreme and are opposed to the schizophrenic and neurotic group means at the other extreme with the depressives occupying an intermediate position. More precisely, the brain-damaged group mean is differentiated, on the one hand, from the normal and neurotic groups' means at a high level of significance ($p <$ .01) by canonical Variate 1, and, on the other hand, from the schizophrenic, neurotic, and depressive groups' means (p again being less than .01) by canonical Variate 2.

This would appear to be of value from a diagnostic viewpoint. Of importance, too, is the fact that whereas the normal *S*s tend to have low scores on the "normality-abnormality" dimension, the situation is the reverse for the schizophrenic group, the means for the respective groups differing in each case beyond the .001 level of significance. Indeed, a careful study of the t test results shown in Table 4 indicates that the canonical variates enable us to distinguish all group means from each other significantly on one or other, or on both, of the variates with the exception only that the neurotic group mean can be differentiated on neither variate from those of the depressive and schizophrenic groups. However, the main purpose of this investigation was to differentiate between the brain-damaged group on the one hand and the remaining groups on the other, and this purpose has been realized.

Table 4

Significance of Differences Between the Groups on the Canonical Variate Scores

F ratios

Canonical Variate I	Canonical Variate II
21.73**	10.226**

t (2-tail)

Groups compared	Canonical Variate I	Canonical Variate II
Brain damaged v. normal	6.321**	1.379
Brain damaged v. depressive	1.648	2.629**
Brain damaged v. schizophrenic	1.868	5.770**
Brain damaged v. neurotic	3.029**	3.425**
Normal v. depressive	3.429**	1.674
Normal v. schizophrenic	3.879**	4.932**
Normal v. neurotic	2.195*	2.482*
Depressive v. schizophrenic	0.001	2.340*
Depressive v. neurotic	1.112	0.569
Schizophrenic v. neurotic	1.222	1.824

* Significance of .05 or above.
** Significance of .01 or above.

Before going on to consider the identification of individual *S*s, we might note that a more meaningful position of the variates might be accomplished by rotation to a position approximately 45 degrees in a clockwise direction. We might then preserve the "normality-abnormality" dimension, and emphasize the interesting relative position occupied by the brain-damaged group. Such a position, possibly reflecting qualitative differences between brain-damaged and other groups, psychiatric and normal, might suggest fruitful lines along which further investigation could be directed.

Differentiating groups, however, is not the same thing as diagnosing individual patients, and so we turn to the second aspect of our study (Figs. 3 and 4) which deals with the problem of classification. Ideally, the canonical variate space should be partitioned into five distinct compartments, one for each group, in such a way that the number of misclassifications is a minimum. This, as noted earlier, can be done efficiently by eye when the space is one of only two dimensions. However, as the means of our neurotic group on the two variates are not differentiated from those of the depressive and schizophrenic groups—though the latter are differentiated from each other on Variate 2—it was decided, in the interest of clarity, to use two figures rather than one for classification purposes, although this is not the ideal procedure.

In the first figure (Fig. 3), the members of the brain-damaged, schizophrenic, and normal groups are represented, and the variate space is partitioned between them so as to achieve a minimum number of misclassifications. A count of the number of individuals correctly and incorrectly classified is given in Table 5. The table shows the percentage of correct classifications to be high, although there is still considerable overlapping between the schizophrenic and brain-damaged groups, 7 members out of the 33 of the latter group falling in the schizophrenic compartment.

In Fig. 4 only the brain-damaged and depressive groups are represented. The results of partitioning the canonical variate space between them are shown in Table 6. Here again, although the percentages of correct classifications are fairly high, 13 out of the 33 brain-damaged *S*s fall in the depressive compartment. Indeed, we must conclude that our efforts at differentiating the brain-damaged patients from the other groups, as seen from both figures, are not quite as successful as one might have hoped. The difficulty is due largely to the fact noted earlier that the variance of the brain-damaged group on the several tests is greater than that of the other groups. This greater variation is reflected in the canonical variate scores with the result that, as seen in Figs. 3 and 4, the brain-damaged patients are widely scattered.

Table 5

Identifications Comparing Normal, Schizophrenic, and Brain-Damaged Groups

	Actual diagnosis		
Test diagnosis	Normal	Schizophrenic	Brain-damaged
Brain-damaged	3	1	23
Schizophrenic	9	21	7
Normal	36	4	3
Total	48	26	33
Percentage correct	75%	80.8%	70%

Table 6

Identifications Comparing Brain-Damaged and Depressive Groups

	Actual diagnosis	
Test diagnosis	Depressive	Brain-damaged
Brain-damaged	4	20
Depressive	13	13
Total	17	33
Percentage correct	76.5%	60.6%

It should be noted that scores on canonical Variate 1 are related to intelligence in a way indicating that high scores on the former are associated with low scores on the latter. As the brain-damaged group were found to be significantly less intelligent than the other groups (2), it is clear that some of the differentiation between this group and others might be a function of differences in intellectual status. Consequently, an analysis of covariance was performed on canonical Variate 1 scores, equating the groups for intelligence level. This resulted in a general slight reduction in significance levels of differences between the groups. This procedure was not necessary in the case of canonical Variate 2 which is not related to intelligence level.

Summary

Four measures of discrepancies in the reproduction of simple designs by drawing were

taken for each individual in five groups: normal, neurotic, depressive, schizophrenic, and brain-damaged. In view of theoretical and practical difficulties encountered in differentiating satisfactorily between these groups on the measures, the technique of discriminant function analysis was applied to the data in the hope that the amount of misclassification would be reduced.

The statistical model was briefly described and also the advantages in using the particular technique chosen.

In terms of levels of significance, the differences between the groups on the canonical variate scores were not strikingly superior to those obtained on the measures before applying the statistical technique. It might be noted that a similar conclusion was reached by Beech (2) after comparing the number of subjects identified as belonging to one or other of the groups under the two conditions. However, it was pointed out that certain distinct advantages were to be gained by employing the discriminant function method to the data, and it was clear that the test measures combined in this way were clinically useful, especially where normals, schizophrenics, and brain-damaged subjects were concerned.

A tentative interpretation of one of the two canonical variates was put forward. The second canonical variate presents a problem in interpretation.

Finally, Vector 1 scores were found to be related to intelligence level, but it was reported that only slight reductions in the significance of differences between the groups resulted when the groups were equated for level of intelligence.

Received May 10, 1957.

References

1. Bartlett, M. S. Internal and external factor analysis. *Brit. J. Psychol.* (Statist. Sec.), 1947–8, **1**, 73–81.
2. Beech, H. R. An investigation of the performance on a perceptual motor task by psychiatric patients, with special reference to brain damage. Unpublished doctoral thesis, Univer. of London, 1956.
3. Eysenck, H. J. Psychiatric diagnosis as a psychological and statistical problem. *Psychol. Rep.*, 1955, **1**, 3–17.
4. Eysenck, S. B. G. A dimensional analysis of mental abnormality. Unpublished doctoral thesis, Univer. of London, 1955.
5. Hamilton, M. The personality of dyspeptics. *Brit. J. med. Psychol.*, 1950, **23**, 182–198.
6. Kendall, M. G. *The advanced theory of statistics.* Vol. II. London: Griffin, 1946.
7. Lubin, A. Linear and non-linear discriminating functions. *Brit. J. Psychol.* (Statist. Sec.), 1950, **3**, 90–103.
8. Maxwell, A. E. The latent roots of nonsymmetric matrices. Unpublished master's thesis, Institute of Psychiatry, London, 1953.
9. Rao, C. R. Utilisation of multiple measurements in problems of biological classifications. *J. roy. statist. Soc.*, 1948, **10**, 159–203.
10. Rao, C. R., & Slater, P. Multivariate analysis applied to differences between neurotic groups. *Brit. J. Psychol.* (Statist. Sec.), 1949, **2**, 17–29.
11. Shapiro, M. B. Experimental studies of a perceptual anomaly. I. Initial experiments. *J. ment. Sci.*, 1951, **97**, 90–110.
12. Shapiro, M. B. Experimental studies of a perceptual anomaly. II. Confirmatory and explanatory experiments. *J. ment. Sci.*, 1952, **98**, 605–617.
13. Shapiro, M. B. Experimental studies of a perceptual anomaly. III. The testing of an explanatory theory. *J. ment. Sci.*, 1953, **99**, 394–409.
14. Yates, A. J. An experimental study of the block design rotation effect, with special reference to brain damage. Unpublished doctoral thesis, Univer. of London, 1954.

The Study

The present study rests on a complete revision of Rapaport's deviant response types. Powers re-evaluated Rapaport's descriptions of his deviant response types, as well as the Rorschach responses which Rapaport lists as examples for each type of deviant response. As a result of this procedure, Powers isolated 10 continua which seemed to be either explicit or implicit in Rapaport's material and which seemed to be relatively distinct and homogeneous.

A short description of each continuum is as follows, with an extreme example of each contained in parentheses following the descriptions:

Continuum I, Cohesiveness—Loosely integrated responses. (Card VIII, W. A room. [What makes it look like this?] Different colored furniture, different shapes.)

Continuum II, Confusion—Any confusion on the part of the subject in communicating his response to the examiner. (Card I, W. Shadow pictures . . . desert picture . . . you lift up your hands in a desert picture . . . shadow what bothers you.)

Continuum III, Form level—The degree of correspondence between the percept and the blot area used. (Card VII, W. It is not a shoelace, is it?)

Continuum IV, Mangling—Mangled or withered content. (Card VII, W. What is left of a bat, mashed up by a truck, part of it decayed.)

Continuum V, Social acceptability—Taboo content, generally of a sexual nature. (Card II, W. Intercourse; glorified kiss.)

Continuum VI, Deterioration color—Gory color responses. (Card IX, W. Not a healthy color . . . looks almost gangrenous.)

Continuum VII, Increase of distance—Any tendency to go beyond the blot with insufficient justification. (Card VI, W, each half. Like going from Helsingfors to Leningrad . . . that kind of country; at dusk; thinly populated . . . lonely feeling.)

Continuum VIII, Loss of distance—Any tendency to be inflexibly and inappropriately bound by the blot configuration. (Card IX, W. Sex organ (upper two-thirds), female . . . these two (lower pink) seem to be men looking at it.)

Continuum IX, Threat—Any response which shows anger, aggression, or fear. (Card VIII, upper middle space. Horrible face . . . fierce . . . twitching lips.)

Continuum X, Affect—Overt expression of positive or negative feelings toward a given card or a given response. (Card IV. A hideous picture.)

Once the 10 continua had been isolated, an analysis of the characteristics of each continuum suggested that they could readily be combined into four large classes. Continua I, II, and III were combined to form Class A, intellectual disorganization; Continua IV, V, and VI were combined to form Class B, deviant content; Continua VII and VIII were combined to form Class C, inappropriate increase or loss of distance; and Continua IX and X were combined to form Class D, affective response.

It was hypothesized that schizophrenics have a significantly higher or more pathological mean score on each of the four classes than either normals or neurotics, and it was also hypothesized that neurotics have a significantly higher mean score on each of the four classes than normals.

No direct comparisons with the studies using Rapaport's deviant response types are possible because of differences in method. Indirect comparisons will be made, however, concerning the respective reliabilities and validities.

Method

Subjects

Seventy-five *S*s were used, 25 normals, 25 neurotics, and 25 schizophrenics. All of the *S*s were white, all were between 20 and 40 years of age, all had at least 100 IQ, and all gave at least 15 Rorschach responses. No *S* was included in the experiment who did not meet all of these qualifications.

The three groups of *S*s were equated for sex, there being 11 males and 14 females in each group. They were also roughly equated for the variables of age, IQ, and education, with the schizophrenic *S*s, as might be expected, being generally older than the normal and neurotic *S*s and also having generally less education and lower IQ ratings.

All of the normal *S*s were selected on the basis of adequate adjustment, with no history of psychiatric treatment and no history of psychiatric disability in any members of the immediate family; all of the neurotic *S*s were out-patients and no *S*s were used who were deemed by the therapist to be a character disorder, to have marked schizoid traits, or whose adjustment to reality was somewhat tenuous; all of the schizophrenic *S*s had been hospitalized for periods ranging from one year through twelve years, with the diagnosis being quite definite for each case.

Table 1

Distribution of Scores for Class A, Intellectual Disorganization

Score	Group		
	Normals	Neurotics	Schizophrenics
10.0	0	0	0
10.0–10.9	18	19	1
11.0–11.9	6	5	5
12.0–12.9	0	0	1
13.0+	1	1	18

Procedure

A scoring scale was drawn up for each of the 10 continua, ranging in five-point intervals from Point 10 to Point 50. Point 10 consisted of an absence of deviance with respect to the characteristics of each continuum, while Point 50 consisted of extreme deviance with respect to the characteristics of each continuum. Point 30 and Point 50 of each continuum were anchored by five illustrative Rorschach responses, taken almost entirely from Rapaport's examples. In addition, a description of the characteristic quality of each continuum was drawn up for the use of the scorers.

One judge scored all 75 Rorschach records. Each response in every record received 10 separate scores, one score for each of the 10 continua. For the purpose of obtaining an estimate of scoring reliability five Rorschach records were chosen at random from each of the three *S* groups and these 15 Rorschach records were scored independently by a second judge following a period of training in the scoring system.

Table 2

Distribution of Scores for Class B, Deviant Content

Score	Group		
	Normals	Neurotics	Schizophrenics
10.0	18	9	10
10.0–10.9	7	15	12
11.0–11.9	0	1	2
12.0–12.9	0	0	0
13.0+	0	0	1

Table 3

Distribution of Scores for Class C, Inappropriate Increase or Loss of Distance

Score	Group		
	Normals	Neurotics	Schizophrenics
10.0	2	0	1
10.0–10.9	16	6	7
11.0–11.9	5	12	7
12.0–12.9	2	6	3
13.0+	0	1	7

The four class scores for each *S* were obtained by averaging the scores of the two or three continua making up each class.

Results

Validity

Tables 1, 2, 3, and 4 contain the distribution of scores for each of the three groups of *S*s for Classes A, B, C, and D respectively. These class scores were obtained by averaging the scores of the constituent continua.

Table 5 gives the *t* ratios obtained in the various intergroup comparisons, with one asterisk indicating significance at the .05 level and two asterisks indicating significance at the .01 level.

The significant differences contained in Table 5 were all in the predicted direction with the exception of the neurotic-schizophrenic comparison for Class D. The neurotic

Table 4

Distribution of Scores for Class D, Affective Response

Score	Group		
	Normals	Neurotics	Schizophrenics
10.0	4	3	14
10.0–10.9	14	10	7
11.0–11.9	4	5	2
12.0–12.9	3	5	1
13.0+	0	2	1

Table 5

t Ratios Based on the Comparison of the Mean Score Values of Each of the Two Classes for 25 Normal, 25 Neurotic, and 25 Schizophrenic Subjects

Groups Compared	Class A	Class B	Class C	Class D
Normal-Schizophrenic	5.59**	1.73	3.54**	1.35
Neurotic-Schizophrenic	5.57**	1.14	1.81	2.35**
Normal-Neurotic	0.10	3.18**	3.76**	1.31

** .01 level.

group had a significantly higher mean score than the schizophrenic group for this class.

It may be seen from Table 5 that two of the four possible comparisons were significant in each of the three intergroup comparisons. Class A and Class C together accounted for four of the six significant differences. It may be seen that Class A significantly differentiated the schizophrenic group from the two nonpsychotic groups, while Class C significantly differentiated the normal group from the two patient groups.

Table 6 contains the distribution of total scores for each group, with the total score for each *S* being obtained by summating the four class scores.

Three intergroup comparisons were made using total scores and all three were significant at the .01 level. The *t* ratios for the normal–schizophrenic, neurotic–schizophrenic, and normal–neurotic comparisons were, respectively, 4.05, 3.27, and 2.95.

Table 6

Distribution of Total Scores, Using the Summation of Scores for All Four Classes for Each Subject

Score	Normals	Neurotics	Schizophrenics
40.0	0	0	0
40.0–41.9	7	3	1
42.0–43.9	14	11	2
44.0–45.9	3	8	5
46.0–47.9	1	2	4
48.0–49.9	0	0	4
50.0+	0	1	9

Table 7

Intercorrelations of the Four Classes ($N = 75$)

Class	B	C	D
A	.65	.65	−.09
B		.69	.03
C			.24

Table 7 contains the intercorrelations of the four classes.

It may be seen from Table 7 that Classes A, B, and C have rather high intercorrelations, while the correlation of each of these three classes with Class D is negligible.

Scoring Reliability

Reliability coefficients were computed for the 10 continua and the four classes. For Continua I through X the reliability coefficients were, respectively, .63, .37, 35, .08, .36, .54, .33, .82, and .76; while for Classes A, B, C, and D the reliability coefficients were .57, .68, .81, and .89.

Since each class was composed of two or three continua, the correlation coefficients of the continua making up each class were averaged in order to find some explanation for the fact that the reliability coefficients for the 10 continua were not more imposing. These correlation coefficients were respectively for Classes A, B, C, and D, .46, .73, .44, and .79. These values are lower than the corresponding reliability coefficients for three of the four classes, and considerably so for Class C. These data, in combination with inspection of the scoring of the records, indicate that the two scorers were fairly well agreed concerning the degree of deviance of responses with respect to each class but that the two scorers were much less in agreement as to which constituent continuum best described the deviance. The two judges tended to disagree to some extent in assigning responses to Continua I or III of Class A and Continua IX or X of Class D.

These differences were not extreme, however, and might well be lessened by a more succinct formulation of these continua. The two scorers, on the other hand, disagreed to a considerable extent in assigning responses to Continua VII or VIII of Class C. It seems clear that these two continua should be combined into one continuum, since duplication existed in the minds of the scorers if not in the descriptions of the continua. Logically, it is a simple matter to separate these two opposed ways of responding to the Rorschach blots. In actuality, however, it is apparently much simpler for a judge to indicate that a given response deviates from the usual mode of responding to the inkblots to a greater or a lesser extent than it is for the judge to characterize this deviation as manifesting primarily "increase of distance" or primarily "loss of distance" from the blots.

Discussion

In the previous studies (1, 3) making use of Rapaport's formulations, the judges were obviously making use of *general* impressions obtained from Rapaport, since they agreed so little in assigning responses to specific scoring categories. The present study was based on the assumption that these general impressions could be isolated from Rapaport's material, and the authors feel that the four classes of the present study represent the core of these impressions.

As stated earlier, no direct comparison is possible with respect to validity and scoring reliability between the present study and the two previous studies which made use of Rapaport's deviant response types. This is so because the present study does not make use of one overall score in determining validity and reliability but, instead, has broken the overall score down into four large classes. We can make a validity comparison on the basis of optimal cutoff points, using the total score from Powers and Hamlin's study (1) and using the single most effective class in each intergroup comparison in the present study. This comparison suggests the two methods are roughly equal in effectiveness in making the same intergroup discriminations. If we consider the distribution of total scores contained in Table 6, the present study is even more effective comparatively, with a cutoff score of 46.0 correctly identifying 18 schizophrenic *S*s and misidentifying, respectively, only one normal *S* and three neurotic *S*s. Even more effective in discriminating between groups, however, is the combination of scores from Classes A and C for each *S*. In the latter instance, a cutoff score of 24.0 correctly identifies 19 schizophrenic *S*s while misidentifying only one normal *S* and two neurotic *S*s. The scoring reliability coefficients in both Watkins and Stauffacher's and Powers and Hamlin's studies were highly significant and quite imposing in size. The scoring reliability coefficients of the four classes used in the present study were also highly significant. Furthermore, there is every reason to expect that these latter values can be increased by better delineation of the continua and the amalgamation of some of the continua. The set of categories developed in the present study seem to have greater potentiality for improvement than do those developed by Rapaport. Moreover, the four classes appear to have more psychological relevance than Rapaport's categories, and these four classes concern four important and relatively nonoverlapping aspects of Rorschach productions. The category of "intellectual disorganization," for example, is much more communicable and self-evident in meaning than are categories such as "fabulized-combination" or "contamination." Such obscure terms as the two latter ones tend to cast a pall of mystery over Rorschach interpretation and tend to isolate the Rorschach "magician" from those of his associates who prefer to deal with more concrete terms. Furthermore, terms such as "intellectual disorganization" are much more amenable to statements concerning degree than are the previously mentioned terms, which almost seem bound to a "sign" approach. May we be delivered from signs and the occult.

Summary

Rapaport's material concerning deviant verbalizations on the Rorschach test was analyzed by one of the authors, and 10 continua were isolated, which seemed to be present explicitly or implicitly in this material. An analysis of the characteristics of these 10

continua showed that they could be combined into four large classes, namely, intellectual disorganization, deviant content, inappropriate increase or loss of distance, and affective response. Scoring scales were devised for each of the 10 continua, and it was hypothesized that the scoring reliability coefficients for each of the 10 continua would prove to be statistically significant, as well as the scoring reliability coefficients for each of the four classes.

It was further hypothesized that schizophrenic *S*s have significantly higher or more pathological scores on each of the four classes than normal *S*s and neurotic *S*s and, also, that neurotics have significantly higher scores on each of the four classes than normal *S*s.

The scoring reliability coefficients of all four classes proved to be significantly reliable, but only five of the ten continua proved to be significantly reliable. In each of the three possible intergroup comparisons, only two of the four classes proved to be significant in each comparison, and one of the two significant differences in the neurotic-schizophrenic comparison was the opposite from that predicted.

Class A, intellectual disorganization, significantly differentiated the schizophrenic group from the two nonschizophrenic groups, while Class C, inappropriate increase or loss of distance, significantly differentiated the normal group from the two patient groups. These two classes were the most effective classes in effecting group differentiations.

No direct comparisons are possible between the results of the present study and the results of the two previous studies making use of Rapaport's categories because of the differences in methods. Indirect comparisons, however, suggest that the present results are at least equal to those of the previous two studies insofar as scoring reliability and intergroup differentiations are concerned. Furthermore, the four classes isolated in the present study appear to have more psychological relevance and are more communicable and less occult than are the previous categories. Future studies are contemplated using the revised versions of the 10 continua.

Received April 17, 1957.

References

1. Powers, W. T., & Hamlin, R. M. Relationship between diagnostic category and deviant verbalizations on the Rorschach. *J. consult. Psychol.*, 1955, **19**, 120–124.
2. Rapaport, D. *Diagnostic psychological testing.* Vol. II. Chicago: Year Book Publishers, 1946.
3. Watkins, J. G., & Stauffacher, J. C. An index of pathological thinking in the Rorschach. *J. proj. Tech.*, 1952, **16**, 276–286.

Journal of Consulting Psychology
Vol. 22, No. 2, 1958

Rorschach Concept Evaluation Test as a Diagnostic Tool[1]

Laverne C. Johnson
Washington University School of Medicine and Research Laboratories of Malcolm Bliss Hospital

Skepticism of the reliability and validity of the clinician's sensitivity in interpreting Rorschach responses and the three to four hours needed to administer, score, and interpret a single protocol has motivated many to search for a quantitative score and a briefer form. The various quantitative and sign scoring systems were reviewed by Knopf (3), and their ability to differentiate effectively among psychoneurotics, psychopaths, and schizophrenics was tested. From his study, Knopf concluded, "For all practical purposes Rorschach summary scores cannot be regarded as effective in differentiating among psychiatric groups" (3, p. 104).

Use of group techniques has fared little better. As the interpretative hypotheses that form the backbone of the Rorschach technique were developed on the basis of an individual test situation, the applicability of these hypotheses to responses obtained in groups has been questioned (2). As a psychometric technique, group methods must rely for their interpretation upon statistically established differentiation among groups. Because of the difficulty in group testing of psychiatric patients, especially schizophrenics, group norms for differential diagnosis are hard to obtain.

The Rorschach Concept Evaluation Technique (CET), though originally developed by McReynolds (4) to study perceptual differences and psychopathology, permits the calculation of the quantitative scores sought by those who have eagerly grasped at the sign approach. Ease in administering the CET and its brevity, about 15 minutes, also offer the possibility of its use as a screening technique in a psychiatric setting and yet it can be individually administered. Essentially, the CET involves the presentation to the patient of 50 specified Rorschach blot areas, each paired with a stated concept. The patient is then asked to decide whether the indicated areas could or could not represent the stated concepts. The patient is requested to answer *Yes* or *No* to each presentation, and his score is determined on the basis of his answers. The test and its development are described in detail elsewhere (5). In his article, McReynolds (5) elaborated on the clinical use of the CET, but made no attempt to evaluate its effectiveness as a diagnostic instrument for individual patients. McReynolds felt at that time that the CET could be best used as an adjunct to the traditional Rorschach technique.

While there is no question that the CET plus the Rorschach will give the most information, the purpose of this study is to investigate the usefulness of the CET independent of the Rorschach. Specifically, how effective is the CET in differentiating among hospitalized psychoneurotics, sociopathic personality disorders, and schizophrenics? In addition, this study contributes data supporting the group norms reported by McReynolds.

Procedure

The patients in this study were all hospitalized at Malcolm Bliss Hospital, a city diagnostic and acute treatment hospital, and carefully examined by four psychiatrists[2] for

---

[1] This study was supported in part by a research grant from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service.

[2] The psychiatrists were George A. Ulett, Eli Robins, Kathleen Smith and Neil McCullough.

Table 1

Mean Age and Mean IQ of Present Sample

Group	*N*	Age		IQ	
		M	*SD*	*M*	*SD*
Paranoid schizophrenic	24	36.79	8.10	105.63	15.00
Schizophrenic other than paranoid	29	38.8	7.16	96.81	14.16
Acute schizophrenic	21	30.31	8.21	102.00	17.65
Chronic schizophrenic	32	33.96	13.22	101.34	13.22
Neurotics	34	31.22	8.67	103.76	13.59
Sociopathic personality disturbance	27	28.35	5.76	103.50	12.52

diagnostic discrimination in a larger study of schizophrenia (6). Three of the four psychiatrists, using stringent written criteria, had to agree as to the diagnosis before the patient was accepted for the project. Testing was done after the patient was accepted and before treatment was begun, usually within one week after admission. The number of patients in each category and the mean ages and IQs are listed in Table 1. The IQ is based upon four subtests of the Wechsler-Bellevue, Form I: Information, Comprehension, Similarities and Vocabulary. There was no significant difference among the groups compared with respect to age or IQ. The mean ages of the sample reported by McReynolds were paranoid schizophrenics, 35.0, schizophrenics other than paranoid, 34.16, and neurotics, 29.16. These ages are not significantly different from those of the present study. McReynolds listed mean education levels and did not report mean IQ scores for his groups.

The CET was administered as part of the research battery and in most cases the regular Rorschach was administered before the CET. However, in some instances, due to pressure of time, the patient was told in the usual instructions to look through the cards, but the responses were not recorded nor was there an inquiry. Comparison of the two approaches revealed no differences in CET scores.

The scores used in this study are those retained by McReynolds for clinical use (5). These scores are *J*, derived from number of *Yes* responses or agreements with the stated concept; *V*, a measure of consistency of judgment which reflects the consistency with which the patient applies his own standards of evaluation (as indicated by *J*) to the various items in terms of the item criteria based upon a normal sample; *E*, derived from the relationship of *J* and *V;* and *R*, number of Rorschach responses, if Rorschach was administered before CET. The *T* score conversion table reported in the original study was used in this study.

Results

In Table 2 are listed the CET scores for McReynold's study and for the present study. As the Rorschach was not administered to all patients, the *N* for *R*, number of Rorschach responses, is smaller than that for scores derived from the CET alone. Comparison of group means in Table 2 reveals a high degree of similarity between the present data and those reported by McReynolds. Inspection of Table 2 also suggests, and *t* tests confirm, that there are no differences between the psychoneurotic and the sociopathic personality trait disorder groups on any of the CET scores. These two groups were then combined and referred to as the nonschizophrenic patients for comparison with schizophrenic patients. The schizophrenic group was categorized as to diagnosis, paranoid and nonparanoid, and duration of illness. Patients whose symptoms were of recent onset, 12 months or less, were called acute. Those ill for more than 12 months were listed as chronic. In Table 3 are the *t* values for the *J*, *V*, *R*, and *E* CET scores with various diagnostic groupings. Values of *t* for *J* and *R* scores do not reach significance for any of the comparisons. The results indicate that neither the number of concepts agreed to on

Table 2

Comparison of McReynold's Normative Data Mean Scores on the CET, with Present Sample

	Concept evaluation test score											
	J			*V*			*R*			*E*		
Group	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
McReynolds												
Paranoid schizophrenics	31	42.72	19.15	28	45.61	9.03	31	45.84	12.30	28	49.78	11.48
Schizophrenics other than paranoid	34	48.21	21.95	28	37.50	13.27	34	43.82	9.72	28	48.50	14.94
Neurotics	34	46.65	9.88	34	50.82	7.64	34	44.06	7.88	34	53.18	8.10
Johnson												
Paranoid schizophrenics	24	37.50	15.19	24	43.00	8.35	17	45.65	12.74	24	51.71	9.56
Schizophrenics other than paranoid	29	45.34	13.56	29	36.41	10.80	21	40.71	12.92	29	43.86	11.61
Neurotics	34	42.70	12.78	34	48.73	8.49	27	44.11	10.28	34	54.54	8.97
Sociopathic personality disturbance	27	43.15	10.36	27	48.52	8.04	21	45.48	10.90	27	51.88	8.32
Acute schizophrenic	21	41.42	14.86	21	34.15	10.42	12	45.58	9.86	21	44.05	9.74
Chronic schizophrenic	32	42.75	15.99	32	42.50	10.46	24	41.50	13.37	32	48.81	13.00

the CET nor the number of responses given on the Rorschach can be used to differentiate among the groups compared. It was expected that the paranoid schizophrenic would agree to fewer concepts on the CET than the other schizophrenic and nonschizophrenic patients. This difference was not statistically significant ($p = .07$).

Table 3

Comparison of CET Scores—*t* Values

	Concept evaluation test score			
Groups compared	*J*	*V*	*R*	*E*
Total schizophrenic—Total nonschizophrenic	.46	4.79***	.60	2.68**
Paranoid schizophrenic—Nonschizophrenic	1.54	2.69**	.33	.66
Schizophrenic other than paranoid—Nonschizophrenic	.71	6.51***	1.20	3.53***
Paranoid schizophrenic—Schizophrenic other than paranoid	1.83	2.34*	1.17	2.52**
Acute schizophrenic—Chronic schizophrenic	.28	2.67**	.98	1.41
Acute schizophrenic—Nonschizophrenic	.42	5.35***	.32	3.45***
Chronic schizophrenic—Nonschizophrenic	.10	2.77**	.95	1.66

* p = .05 level.
** p = .01 level.
*** p = .001 level.

Consistency of judgment score, *V*, is the only CET measure that is significantly different for all the groups compared. This difference for schizophrenics and nonschizophrenics is highly significant. If a *V* cutting score of 40 or below is used, only 9 (14.7%) of the nonschizophrenic patients receive scores below 40, while 31 (65%) of the schizophrenic groups receive scores below 40. Most of the schizophrenics who receive *V* scores above 40 are in the chronic group. Eighteen (56.2%) of the chronic group receive *V* scores above 40 while only 4 (19%) of the acute cases receive *V* scores in the normal range. Thus, the *V* score seems to be a fairly sensitive indicator of an inconsistency of judgment which is seldom seen in nonschizophrenics, and less often in chronic schizophrenics, but which is fairly characteristic of the acute schizophrenic patient.

McReynolds (5) speculated that the variability of standards implied by a low *V* score may be a function of one of two factors. First, the low *V* may indicate a fluctuating, unpredictable set of standards. Second, the low *V* may indicate that the patient's standards of evaluation are partly a function of the meaning to him of the stimuli. For example, a patient might tend to say *No* when the concept presented referred to a personality area against which he had a need to defend. While one or both of the above may be true in some patients, behaviorally *V* seems to be most related to the effectiveness of the patient's controls at the time of testing. Also *V* may be used to evaluate changes induced by therapy.[3]

E, *J* minus *V*, was the least defined of the clinical scores by McReynolds. It was tentatively conceived as a measure of rigidity, but Block's work (1) offered no strong support for this view. The results of this study also do not offer support for *E* as an important variable. While the difference between schizophrenics and nonschizophrenics was significant, if the cutting score of 40 or below is used again 38 (71.3%) of the schizophrenic group receive "normal scores." However, only three (4.8%) nonschizophrenic patients received an *E* score below 40. Again like *V*, the *E* score seems to be highly diagnostic if below 40 but not too helpful if above 40. No particular pattern of low *E* scores seemed to be present except that patients who agree to a large number of concepts have loose standards of judgment and are more likely to have a low *V* and thus a low *E*. This seems to be a consistent pattern which has been referred to above as indicative of inadequate controls. Thus the *E* score adds little to the *J* or *V* score interpretations.

While no systematic study of mental defectives and organics was made because of the small number of cases seen ($N = 12$), the scores from these groups, especially the *V* score, were similar to the schizophrenic group. Chronic brain syndromes (paresis and alcoholism) and patients with IQ below 60 tended to have *V* scores below 40. These findings agree with those reported by McReynolds (5).

Summary and Conclusions

The Rorschach Concept Evaluation test has been investigated as a technique for obtaining a quantitative score that might be useful in diagnosis and as a brief screening technique which could be individually administered. In the present study, data of 63 carefully diagnosed schizophrenics and 34 neurotics were consistent with the original norms reported by McReynolds for schizophrenics and neurotics. The effectiveness of the CET as a diagnostic tool independent of the Rorschach was evaluated by comparing the clinical scores *J*, *V*, *E*, and *R* received by schizophrenics, neurotics, and sociopathic personality disorders. No difference between the neurotics and the sociopathic group was found. Only the *V* score significantly differentiated between the schizophrenic and nonschizophrenic groups. Only 15% of the nonschizophrenic group received scores indicating marked inconsistency of judgment, while 65% of the schizophrenic group received scores indicative of inconsistent standards. Acute schizophrenics had lower *V* scores than chronic schizophrenics and the *V* scores of paranoid schizophrenics were not as low as other schizophrenics. The *V* score is felt to be useful in diagnosis, but its use as an indicator of the patient's control at the time of testing regardless of diagnosis may be more important. The *E* score added little information beyond that given by the *J* and *V* scores.

---

[3] Twelve patients on whom records were secured six months after admission showed a significant increase in mean *V* scores, 36.8 to 48.7. Four patients initially had *V* scores above 40, while three had *V* scores of 40 or below at six months. These three were the only patients still in the hospital at time of retest.

Received May 24, 1957.

References

1. Block, J. An experimental investigation of the construction of ego control. Unpublished doctoral dissertation, Stanford Univer., 1950.
2. Klopfer, B., Ainsworth, Mary D., Klopfer, W. G., & Holt, R. R. *Developments in the Rorschach technique.* Vol. 1. Yonkers, N. Y.: World Book, 1954.
3. Knopf, I. J. Rorschach summary scores in differential diagnosis. *J. consult. Psychol.,* 1956, 20, 99–104.
4. McReynolds, P. Perception of Rorschach concepts as related to personality deviations. *J. abnorm. soc. Psychol.,* 1951, 46, 131–141.
5. McReynolds, P. The Rorschach concept evaluation technique. *J. proj. Tech.,* 1954, 18, 60–74.
6. Smith, Kathleen, Ulett, G. A., & Johnson, L. C. Convulsive threshold in schizophrenia. *Arch. Neurol. Psychiat.,* 1957, 77, 528–532.

Journal of Consulting Psychology
Vol. 22, No. 2, 1958

Movement Responses and Creativity[1]

Dorothy Park Griffin
North Carolina State Board of Public Welfare

This study deals with the tendency of human beings to perceive movement in suitable stimuli presented pictorially in a static visual field. The well-known *M* response of the Rorschach technique is the typical example of such a movement response. The behavioral correlates of *M*, according to Rorschach, are "more individualized intelligence, greater creative ability, more 'inner life,' stable affective reactions, less adaptable to reality, more intensive than extensive rapport" (5). The purpose of this study is to investigate the relationship between movement responses and these behavioral correlates, particularly creativity.

Beck, Hertz, also Klopfer and Kelley have affirmed agreement with Rorschach's statement of these relationships. Schachtel (7) is somewhat more cautious, stating that *M* responses "represent a factor in the capacity for creative experience." However, Anne Roe (4), in her research on eminent painters, found no clear relationship between *M* and creativity. In fact, she found a relatively low number of *M* responses. Rust (6) found a slight but significant negative correlation between creativity and the movement response as elicited by the Levy Movement Blots. Zubin, referring to seven studies of creative ability conducted in the laboratory of the New York Psychiatric Institute on creative versus noncreative writers, mathematical statisticians, and high school students, states that "all have failed to reveal any differences on Rorschach performance, and even tests especially designed to elicit movement have failed" (9).

Procedure

Using the Levy Movement Blots, a technique devised by David Levy, with the express purpose of eliciting movement, two experiments were conducted with students in a women's liberal arts college to investigate the relationship between movement responses and creativity.

Selected for the first study were 20 college women rated as highly creative by at least one teacher and two students and 20 others, matched in so far as possible in age, sex, year in college, and intelligence (measured by ACE scores), who were rated by at least one teacher and two students as noncreative. An effort was made to secure a sampling of creative students from all major departments. However, the majority of the group judged as highly creative turned out to be majors in art, English, and home economics, while the noncreative group included no art or home economics majors and only two English majors.

No formal rating scale of creativity was used, but heads of major departments were asked to select students judged to be genuinely creative, according to their own standards and also in the light of Murphy's definition (3). Then they were asked to name several more whom they judged to be average in general ability but lacking in creativity. Student raters were given the same instructions.

The mean age of the creative group was 21.5 years and of the noncreative 21.2, range 20–22 years. Sixteen of each group of 20 were seniors in college and four were juniors. The mean ACE score for the creatives was 99.75; range 58–134. The mean ACE score for the noncreatives was 98.5; range 63–141.

Since several other women judged as highly creative could not be included in the carefully equated group because of highly superior ACE scores, it was decided to run a second experiment in which the sample would be composed of the first two groups plus five additional highly creative subjects with superior ACE scores and five additional students judged as noncreative, matched approximately for age and year in college but not for intelligence, since such matches seemed to be unavailable. In this way it was thought that something might be learned of the extent to which intelligence, as well as creativity, plays a part in the movement response. The mean ACE score for the five additional creatives was 132.6 as compared with 91.6 for the five additional noncreatives.

---

[1] This research was done while the author was at Meredith College, Raleigh, North Carolina.

The Levy Movement Blots were administered individually to each of the 50 women in the two studies, according to the directions in the Zubin Manual (8) and scored on the basis of 21 scales, including Levy's three-step evaluation scale, the Zubin scales, and three scales devised by the author to measure Part B of the Levy, called the "phantasy test." The examiners were seven psychology majors trained by the author. Each of the examiners first took the test herself and then spent several hours in group sessions with the author, scoring responses and discussing scoring problems. Later sessions were devoted to the recording and group scoring of a demonstration test administered by the author. No statistical analysis has been made for examiner errors, but the group had uniform training with ample opportunity for discussion of problems and for checking with other examiners and with the author on doubtful points. Further to minimize examiner errors, such as the halo effect, an effort was made to divide, as equally as possible, the number of creative and noncreative students tested by each examiner. Complete protocols of each subject were obtained for Parts A and B of the test, typed and scored by the student examiner, and spot-checked by the author.

Results

The mean scores and differences, the standard deviations and critical ratios on the 21 movement scales were computed for the creative and noncreative groups in the two studies. Only one scale, "Control of Movement," showed a difference significant beyond the five per cent level of confidence. On both studies this can best be interpreted as entirely negative. The method of "rank scores" as well as the t test was used to measure the comparisons. The two tests yielded the same results, even the probability levels being very close. The data were also analyzed by the chi-square method, which gave substantially the same results.

Further to test the relationship of intelligence to differences on the movement scales between creative and noncreative groups, the ACE scores of each group of 25 were ranked and divided into high, medium, and low intelligence subgroups. The range of the creative subgroups was: high, 119–162; medium, 91–115; low, 58–89; and of the noncreative subgroups: high, 110–141; medium, 86–110; low, 63–84. The means of each subgroup of creatives and noncreatives on the Control of Movement scale, on which a significant negative difference had been obtained, were compared. The difference in each case was in favor of the noncreative group, i.e., negative.

While the difference in ACE level was definitely greatest in the high group of the creatives, the significant differences between creatives and noncreatives, on this scale, is found not at this level but between the medium intelligence groups, where the ACE scores vary only a few points both in means and range. Thus it would seem that the significant difference is quite independent of the intelligence factor.

A comparison of the mean scores and differences of the group of 25 creatives on the 21 scales was also made with a norm group of 100 women from the same college, selected at random from the various college classes and given the test during the following year by ten psychology majors, trained by the author, as previously described. No significant differences were found, but the direction of difference was negative as in the other studies.

Discussion

In general, these negative findings agree with those of Roe (4), Rust (6), and Zubin (9). The weight of experimental evidence seems to be definitely against the classic Rorschach interpretation of M as signifying "more individualized intelligence" and greater creative ability. It becomes increasingly evident, as Burchard states, "that we cannot reach the secret of creativity by counting *M*'s" (1). In fact, Rorschach himself regarded his work as experimental and suggested that his conclusions be regarded more as observations than as theoretical deductions, since, as he wrote in 1911, "the theoretical foundation for the experiment is for the most part still quite incomplete," a statement which could well describe the situation today (2).

Received May 15, 1957.

References

1. Burchard, E. M. L. The use of projective techniques in the analysis of creativity. *J. proj. Tech.*, 1952, **16**, 412–427.
2. Klopfer, B., Ainsworth, Mary D., Klopfer, Walter G., & Holt, Robert R. *Developments in the Rorschach technique*. Vol. I. Yonkers, N. Y.: World Book, 1954.
3. Murphy, G. *Personality*. New York: Harper, 1947.
4. Roe, A. Artists and their work. *J. Pers.*, 1946, **15**, 1–40.
5. Rorschach, H. *Psychodiagnostics*. New York: Grune & Stratton, 1942.
6. Rust, R. M. Some correlates of the movement response. *J. Pers.*, 1948, **4**, 369–401.
7. Schachtel, E. G. Projection and its relation to character attitudes and creativity in the kinesthetic response. *Psychiatry*, 1950, **13**, 69–100.
8. Zubin, J., & Young, K. M. *Manual of projective and cognate techniques*. Madison, Wis.: College Typing Co., 1948.
9. Zubin, J. Failures of the Rorschach technique. *J. proj. Tech.*, 1954, **18**, 303–315.

Journal of Consulting Psychology
Vol. 22, No. 2, 1958

Approaches to Reliability of Projective Tests with Special Reference to the Blacky Pictures Test[1, 2]

Samuel Granick and Norma A. Scheflen
St. Christopher's Hospital, Philadelphia, Penna.

Reliability data are frequently sparse or absent for projective tests. Twenty tests are reviewed in Anderson and Anderson (2), but significant information on reliability is presented for only four tests, preliminary findings for five, and no information for the remainder. Even where reliability has been reported in detail, incompleteness and uncertainty remain.

Reviews by Macfarlane and Tuddenham (6) and by Ainsworth (1) point to a variety of difficulties in treating projective test data by statistical methods. One difficulty is the development of objective scores which reflect the integrity or Gestalt of the personality. Without such scores subjective judgment must be resorted to, which may confound test reliability with judge reliability.

A more important problem, perhaps, is the relationship between test reliability and the stability of the personality. Since behavior is not usually consistent, marked differences in responses might be expected on repeated administrations of a test. The personality is assumed, however, to have a basic organization which should be reflected in the test responses. Reliability, therefore, may be expected when the clinical features are used as a basis for measurement.

The various attempts to study reliability of projective tests seem to direct attention to these problems rather than to solve them. Since validity implies reliability, the suggestion has been made to emphasize the validity of the test, thus, in effect, bypassing the problem (6). Another suggestion is to determine the limits within which the test is reliable by delineating the aspects of the test data which are unreliable because of such factors as the testing environment and the mental set of the subject (1).

The present authors feel that a direct attack on the problem of reliability is possible. It is suggested that aspects of reliability which involve clinical features of the test be sought. The demonstration that many of these elements are reliable would make it reasonable to place confidence in the test's stability, at least, within a clinical framework.

The above approach is illustrated in a series of reliability studies conducted as part of a more general research on the usefulness of the Blacky Pictures Test for children. It was found possible to make use of three traditional measures of or approaches to reliability. These were (*a*) judgment reliability or the extent to which the scorers agree, (*b*) temporal reliability or the extent to which the test-retest scores are consistent, and (*c*) internal reliability or the extent to which split halves are consistent.

Subjects

The *S*s were grade school pupils, 28 girls and 12 boys, whose ages ranged from six years, six months to eleven years, three months. Their IQs on the Revised Stanford Binet, Form L, ranged from 81 to 127. All *S*s had been evaluated as being well-adjusted by

---

[1] This study was supported in part by a grant from the United States Public Health Service, Institute of Mental Health. The authors gratefully acknowledge the critical and helpful assistance of the psychology staff of St. Christopher's Hospital, Philadelphia, Penna. Sincere thanks also go to the Board of Education, the Hunter, the Miller, and the McKinley Schools of Philadelphia for helpful cooperation.

[2] Presented at the Eastern Psychological Association meeting at New York, April 12, 1957.

Table 1

Percentage Agreement of Judges in Rating Spontaneous Story (Strong or Weak) of Each Card ($N = 40$)

Card No.	% Agreement	*p* Value (Binomial)
I	75	.01
II	58	.34
III	90	.001
IV	70	.01
V	83	.001
VIA[a]	83	.001
VIB	80	.001
VII	68	.05
VIII	73	.01
IX	60	.21
X	95	.001
XI	95	.001

[a] Card VI was scored twice, A for "castration anxiety," B for "penis envy."

means of teacher ratings and clinical judgments prior to the administration of the Blacky Pictures Test. All testings, including the administration of the Blacky, were on an individual basis. Twenty *Ss* (Group A) had a second administration of the Blacky three months to one year later. The other 20 *Ss* (Group B) were selected from a larger group[3] and were individually matched to Group A for age, sex, IQ, and socioeconomic level. The mean age for Group A was nine years, five months and for Group B, nine years, six months. The mean IQ for Group A was 106.8 and for Group B, 107.5. The difference between the means is not significant.

Judgment Reliability

Reliability of scores is basic to the use of any test. With projectives, there has frequently been the problem of devising a scoring system which abstracts the data in a clinically meaningful and reliable fashion. Since totally objective scoring generally proves difficult or impossible to devise, reliability of scoring, in effect, is dependent on the reliability of judgments. In the case of the Blacky, Blum (3) suggests a rating of "strong" or "weak" as one type of score for the emotions elicited by each of the pictures or personality dimensions alleged to be tapped by the test. Accordingly, the following hypothesis suggests itself. Judges agree on the scoring of the spontaneous stories of the Blacky Pictures Test.

Using the criteria for scoring "strong" or "weak" proposed by Blum, with modifications by Hilgeman (5), Winters (7), and the writers, each of 10 judges rated the stories of two cards for the 40 *Ss* of Groups A and B. The validity or clinical significance of scoring was not in question. Merely considered was the extent to which the test stories could be scored by this system in a consistent fashion.

Table 1 shows the percentage of agreement of the judges for each card. That agreement is better than chance is supported at less than the .05 level of confidence for all but two of the cards, the majority of *p* values being .01 or less. It should be noted, however, that only three of the eleven groups of judgments show agreement of 90 per cent or over. The results suggest that consistency of scoring may be achieved but that improved concreteness of scoring criteria needs to be developed.

Temporal Reliability

As suggested earlier, repeated administrations of a projective test to the same *S* over a period of time are not expected to produce the same responses, but they should reflect the stable components of the personality. Despite the undefined nature of the components,

Table 2

Percentage of Correct Matching Between Two Test Administrations to Same *Ss*

Judge	*N*	% Matching	*p* Value[a]
1	6	100	.001
2	6	100	.001
3	7	43	.07
4	7	14	.37
5	7	100	.001
6	7	72	.001
7	7	72	.001
8	7	57	.02

[a] This value was determined by a direct application of probability theory as suggested by Feller (4, pp. 62–66).

---

[3] This larger group consisted of 88 *Ss*, selected from the Philadelphia public schools, to be used in a series of studies of the Blacky Pictures Test. Each child was considered well-adjusted as rated by his teacher and was further judged well-adjusted by a clinical psychologist before the Blacky was administered.

Table 3

Judge Agreement in Rating Similarity of Thematic Content for Two Test Administrations to Same *S*s and for Single Administrations to Matched *S*s ($N = 40$)

Card No.	% Agreement	p Value (Binomial)
I	70	.02
II	73	.01
III	68	.05
IV	60	.21
V	78	.001
VI	75	.01
VII	65	.06
VIII	75	.01
IX	55	.52
X	88	.001
XI	83	.001

reflections or indirect indicators may be used to evaluate the reliability of a test. In the present study, three hypotheses were tested as an approach to the clarification of this principle.

Hypothesis 1. Blacky Test records obtained at different times with the same *S*s can be matched. The Blacky protocols of the 20 *S*s of Group A who had been tested twice over a period of three to twelve months were divided into three sets, one of six boys and two of seven girls. Two judges matched the set of six, and three judges matched each of the sets of seven. Matching was done on a gross basis, permitting the utilization of a variety of clues such as language, story content, and emotional patterns. The judges were informed, furthermore, of the sex of the *S*s.

All matchings but one were beyond chance expectation as shown in Table 2. Perfect matching was obtained in only three of the seven matchings. It appears, however, that there was sufficient temporal consistency in the *S*s' responses generally to enable five of the eight judges to recognize which records belonged together.

Hypothesis 2. Thematic content of spontaneous responses to each of the Blacky Pictures is more frequently similar for the same *S*s on repeated test administrations than for a group of *S*s and their matched controls. The spontaneous stories from the two sets of Blacky protocols of Group A and those obtained from Group B were used to test this hypothesis. Two judges evaluated the thematic content of the three spontaneous stories presented for each of the 11 Blacky cards. Judgments were then made as to whether the stories or responses were similar to or different from each other. For the guidance of the judges, it was suggested that they consider the manifest rather than the implied content in determining whether any two stories were similar to each other.

Agreement between judges is presented in Table 3. There would seem to be only fair consistency between the judges, but yet the overall pattern is one of greater than chance agreement. To test the hypothesis, a paired t was computed for the differences between the number of stories judged similar (t equals 7.25). The hypothesis is supported at the .001 level of confidence.

The finding points to the overall trend in the direction of significantly greater similarity of thematic productions between the two administrations of the test to Group A than between the single administrations to Groups A and B. Despite the much better than chance

Table 4

Percentage of Similarity of Thematic Content Between Two Test Administrations to Same *S*s and Between Single Administrations to Matched *S*s

*S*s Pair No.	Same *S*s	Matched *S*s
1	55	32
2	50	23
3	77	36
4	86	32
5	55	41
6	45	50
7	36	18
8	82	27
9	59	36
10	77	41
11	77	27
12	73	59
13	63	50
14	50	41
15	50	5
16	50	27
17	50	32
18	77	41
19	68	41
20	23	18

Table 5

Consistency of Likes and Dislikes of Particular Cards for Two Administrations to Same *S*s ($N = 20$)

Card No.	Same pref.	*p* Value (Binomial)
I	15	.001
II	15	.001
III	14	.001
IV	14	.001
V	17	.001
VI	19	.001
VII	15	.001
VIII	14	.001
IX	16	.001
X	14	.001
XI	17	.001

result, it should be noted that the amount of consistency in thematic productions by the individual *S* is not very high, as shown in Table 4.

*Hypothesis 3. S*s are consistent in their likes and dislikes for individual cards over a period of time. The preferences or "like" and "dislike" choices for the two sets of Blacky protocols of Group A were compared. Table 5 shows the extent to which each *S* placed the individual cards in the same category for the two administrations of the test. Marked consistency is evident in the choices, supporting the hypothesis at the .001 level of confidence for each card.

Internal Reliability

The writers recognize that a split-half approach to reliability tends to destroy the global quality of a projective test. Nevertheless, it seems feasible to consider the data in terms of specific, clinically meaningful dimensions which can be evaluated through a sampling of the test responses of the individual *S*s. Such an approach serves to maintain the overall unity or Gestalt of the test record and yet permits its division into two approximately similar halves. Along these lines the writers attempted to study the test's reliability with regard to (*a*) verbal fluency as measured by the number of words in the spontaneous stories, and (*b*) organization as reflected through the use of the central theme of each picture.

Hypothesis 1. The verbal fluency of *S*s is consistent within the Blacky Test.

The number of words for each spontaneous story of the 40 *S*s (the first test administration of Group A and the records of Group B) was counted. The total of each *S* on Cards 2, 4, 6, 8, and 10 was compared with his total for Cards 3, 5, 7, 9, and 11, Card 1 being omitted because of the uncertain effect of its unique position in the series. The rank-order correlation between the two halves was .92, which is significant at the .01 level of confidence. The similarity of fluency between the two halves of the test is also reflected in the means and standard deviations (mean of 2–10 equals 184 words and *SD* equals 107; mean of 3–11 equals 178 words and *SD* equals 129; *t* equals 0.64).

Hypothesis 2. In organizing their spontaneous stories, the central theme of each card will be used in a consistent fashion by the *S*s.

The same division of the test as described above was used. Two judges scored each story of the 40 records as to whether it was "structured" (utilized the central theme of the picture) or "unstructured" (failed to use the central theme of the picture). Agreement between the judges was 100 per cent. The phi coefficient of correlation between the two halves of the test was found to be .67, which is significant as the .01 level of confidence.

Discussion and Conclusions

The results reported do not represent high level reliabilities, but since they are uniformly in the same direction they command some respect. Accordingly, the Blacky Pictures Test would seem to have a significant degree of reliability when used with a group of children.[4] Of perhaps greater interest, however, are the implications derived from the approaches used for the study of reliability. The concept of a continuing study of a test's reliability is suggested. What seems required is that many aspects of the test should be investigated over a period of time and under varying circumstances. In this fashion, there is a gradual accumulation of data clarifying the nature of the test's stability. Since an overall "coefficient of reliability" for a projective test does not seem to be feasible, this approach might represent a reasonable alternative.

---

[4] Since there is a five-year age span in the group, and test responses are expected to differ with age, the question arises as to the effect of the age factor on the reliability results obtained. The small sample and design of the study, unfortunately, did not permit an adequate control for age. Inspection of the data, however, reveals no marked differences in consistency of response patterns between the older and the younger *S*s. This, of course, is not a definitive answer to the problem. Further study of the age factor is in order.

From a clinical standpoint, there are advantages to such an approach. Data are provided which throw light on the significance of test results obtained under varied circumstances. Thus, the objectivity of interpretations and judgments is improved. Since hypotheses, such as those tested in this study, tend to be related to qualities of the test generally given attention in the clinical situation, the clinician can more easily differentiate the stable aspects of the test performances from those which are a function of the immediate situation.

In all, the present study attempts to suggest that the traditional concepts of reliability need not be abandoned in dealing with projective materials. Some broadening of the interpretation of the concepts is called for. Also necessary is the adaptation of the test data to available statistical or analytical techniques, but with care being exercised to preserve the clinical entity of the responses being studied.

Summary

In this study, the feasibility of developing reliability measures of projective tests based on the clinical aspects of the test material is considered. Using data obtained with the Blacky Pictures Test on 40 school-age children, several hypotheses are explored related to judgment, temporal, and split-half reliabilities.

The data are analyzed with a view to maintaining the global or clinical quality of the test responses. Evidence is derived which supports the test's stability to a modest degree.

Attention is given, in addition, to the difficulties of deriving an over-all "coefficient of reliability" for projective measures. This study indicates that integration of varied approaches to a test's consistency may serve as an appropriate alternative.

Received May 6, 1957.

References

1. Ainsworth, Mary D. Problems of validation. In B. Klopfer, Mary D. Ainsworth, W. G. Klopfer, & R. B. Holt (Eds.), *Developments in the Rorschach technique*. Vol. 1. Yonkers, N. Y.: World Book, 1954.
2. Anderson, H. H., & Anderson, Gladys L. *An introduction to projective techniques*. New York: Prentice-Hall, 1951.
3. Blum, G. S. A study of the psychoanalytic theory of psychosexual development. *Genet. Psychol. Monogr.*, 1949, 39, 3–99.
4. Feller, W. *An introduction to probability theory and its application*. New York: Wiley, 1950.
5. Hilgeman, Lois M. Developmental sex variations in the Blacky test. Unpublished doctoral dissertation, Ohio State Univer., 1951.
6. Macfarlane, Jean W., & Tuddenham, R. D. Problems in the validation of projective techniques. In H. H. Anderson & Gladys L. Anderson (Eds.), *An introduction to projective techniques*. New York: Prentice-Hall, 1951.
7. Winter, Louise M. Development of a scoring system for the children's form of the Blacky pictures. Unpublished doctoral dissertation, Univer. of Michigan, 1954.

Journal of Consulting Psychology
Vol. 22, No. 2, 1958

Personality Correlates of Q−L Differentials on the ACE[1, 2]

B. Spilka
University of Denver

and Gloria Kimble
Washburn University

In 1952, Altus (1) empirically attempted to determine correlates of intellectual variation in a population of women students. Because of the empirical nature of this work, the authors felt it desirable to verify and explain Altus' findings further.

Altus item-analyzed certain MMPI items against ACE Q − L differentials. Of the 566 items available, 43 met his criterion of selection, but only 26 attained statistical significance at the .05 level.

Nine categories of 3 to 8 items were then constructed, and descriptions of the subjects were based on tendencies to answer these items in a certain direction. The validity of such item groups and this type of interpretive procedure is questionable.

Recognizing some of these difficulties, Altus repeated his study and found that a validation correlation between the Q − L scores and the items dropped from .63 to .25. However, the latter coefficient was significant and thus assumed to indicate validity for the items.

In the present study, 87 women students at Washburn University served as *S*s. Each was administered the ACE, the complete MMPI, the Thurstone Scale of Religious Orthodoxy, the Taylor Manifest Anxiety scale, the F scale, the Siegel Manifest Hostility scale, and measures of Behavioral, Feeling-Tone, and Somatic manifest anxiety developed by Spilka and Siegel.

The MMPI *Mf* scale was used to assess Altus' categories of sexual inversion and immaturity, and masculine-feminine vocational preferences. The Orthodoxy scale, the F scale, and the MMPI *Pd* scale appeared to correspond closely to Altus' "religiose" and "rose-colored glasses" items. The MMPI *Pt* scale was considered a measure of the *S*s' obsessive-compulsive tendencies.

The MMPI *Hs, D,* and *Hy* scales, plus the 4 scales of manifest anxiety, were used to assess Altus' "anxiety" syndrome. The "social sensitivity" category was measured by the MMPI *Si* scale and the Behavioral Anxiety scale. Altus' "resentful attitudes" category was assessed by the MMPI *Pd* scale and the Manifest Hostility scale.

Only 6 of Altus' items now satisfied his acceptance criterion and only 2 were significant at the .05 level. Neither of these items satisfied the acceptance criterion, and only one of them was significant in Altus' study. None of the correlations computed to assess the validity of Altus' categories attained significance.

It was found that 30 of the 43 items exhibited scoring tendencies in agreement with those obtained by Altus. The correlation between the 43 items and the differentials was .22, revealing significance at the .05 level, and supporting Altus' "modicum of validity" claim. Where it lies and what comprises it remains unknown.

It is recommended that further work in this area be conducted on a theoretical rather than an empirical basis.

Brief Report.
Received November 19, 1957.

Reference

1. Altus, W. D. Personality correlates of Q − L variability on the ACE. *J. consult. Psychol.,* 1952, **16,** 284–291.

---

[1] The authors would like to express their thanks to E. D. Turner for his statistical help.

[2] An extended report of this study may be obtained without charge from Bernard Spilka, Department of Psychology, University of Denver, Denver 10, Colorado, or for a fee from the American Documentation Institute. Order Document No. 5468, remitting $1.75 for microfilm or $2.50 for photocopies.

Journal of Consulting Psychology
Vol. 22, No. 2, 1958

Learning in Aphasic Patients[1, 2]

Leo Katz
VA Outpatient Clinic, Brooklyn, N. Y.

Aphasic disturbances are generally defined as impairments of language functions due to brain damage, either through cerebro-vascular accident or external injury to the brain. The extent to which other mental functions are impaired in aphasia is a subject of considerable controversy.

This study attempted to investigate the presence of certain learning deficiencies in aphasic patients often described by aphasia therapists (6, 8, 10). These workers report that frequently a patient may be unable to repeat a phrase or word he has just heard when he is asked for it, but that sometimes, when he is not asked for it, or when he is permitted or encouraged to use roundabout methods, he may be able to furnish the material previously learned. Similar observations were reported by Conrad (1), who studied 24 cases of predominantly expressive aphasics who had word-finding difficulties. He explained their impairment as an inability to structure the conceptual field, so that peripheral fragments from the background of the Gestalt are offered by the patient as his best approximation to the "final" and inaccessible figure, the word itself. To facilitate recall in such patients, he advocated that they be encouraged to supply whatever peripheral or tangential elements are available to them in their efforts to arrive at the desired solution. To some extent, this roundabout method of recapturing lost or forgotten material is also used by non-brain-injured people when trying to remember names or dates (3).

Thus, under certain conditions, making an effort or concentrating on a task does not seem to result in more efficient performance than is obtained in the absence of such self-direction. The present investigation was undertaken to determine whether aphasics exhibit the usual gain in efficiency of recall which results from the intention to remember as compared with the absence of such an intention.

Traditionally, learning has been found to be better with goal-directed than with incidental instructions. It seems obvious that a person making an effort to learn will perform more adequately than one not making such an effort. The hypothesis of the present study, accordingly, is that goal-directed learning, as compared with incidental learning, is significantly more efficient for nonaphasics than for aphasic patients.

Method

The material to be learned consisted of two series of 20 postage stamps (Series X and Y), which were exposed to the subjects (*S*s), one stamp at a time for five seconds each. In the incidental learning situation, one series of stamps was exposed with instructions to evaluate the stamps aesthetically ("tell me whether you like it or not"). After all of the 20 stamps had been shown, they were mixed with a group of 40 other stamps which the *S* had not been shown and he was asked to pick out all the stamps he could remember. The learning score was the number of stamps correctly selected minus one-half the number of stamps incorrectly selected.

In the goal-directed learning situation, which always followed the incidental learn-

---

[1] This paper is based on a dissertation submitted in candidacy for the degree of doctor of philosophy at Teachers College, Columbia University. The writer is indebted to members of his thesis committee, E. J. Shoben, Jr., P. E. Eiserer, Virginia M. Axline, and H. Solomon for their suggestions and encouragement.

[2] Presented to the twenty-sixth annual meeting of the Eastern Psychological Association, Philadelphia, Pa., 1955.

ing situation by about 24 hours, the *S*s were shown the other series of postage stamps, one at a time for five seconds each, this time with instructions to remember them so as to be able to select them afterwards from another group of 40 stamps. The learning score again was calculated in the manner described previously.

The difficulty of the task depended in part upon the degree of similarity between the 20 stamps to be remembered and the 40 stamps from which the 20 had to be selected. Attempts were made to equate this degree of similarity in both series. In order to determine the equivalence of the two series, they were administered to a normative group of 18 *S*s with goal-directed instructions only. The means and standard deviations of the scores obtained under these conditions were essentially equivalent. The split-half reliabilities of the scores of the normative group were $r = .85$ for Series X and $r = .89$ for Series Y, after the Spearman-Brown correction.

Stamp Series X and Y were alternately administered to successive *S*s, e.g., *S* #1 might have Series X administered with incidental instructions and Subject #2 would then have Series Y administered with incidental instructions. Thus, any difference obtained with goal-directed and incidental instructions were probably not due to differences in difficulty between the two stamp series.

Subjects

The *S*s for this study were selected from two Veterans Administration hospitals and one clinic which cooperated in this research.[3] Stamp collectors were excluded from the study.

The experimental group consisted of 25 native white male patients between the ages of 19 and 51 who were being treated for aphasia and right hemiplegia due to either cerebrovascular accident or traumatic injury to the brain. All of the aphasics were being given language retraining and had been classified by the language therapists as predominantly expressive. No acutely aphasic patients were included in the group.

The control group consisted of 25 native, white male patients between the ages of 19 and 48 from the surgical wards of a Veterans Administration hospital. Patients with histories of psychiatric conditions, gastric or duodenal ulcer, or essential hypertension were excluded. The patients were tested postoperatively and each name was first submitted to the physician in charge in order to eliminate those with postoperative complications.

A comparison of the mean age and education of the *S*s indicated that the two groups were essentially homogeneous with respect to age and education. The two groups were found also not to differ significantly in sociocultural status, as measured by occupational level with the Goodenough Anderson Occupational Rating Scale (5).

Results

The hypothesis states that goal-directed learning as compared with incidental learning is more efficient for nonaphasics than for aphasics. To test this hypothesis, the mean recall scores for each type of learning were compared within each of the two groups.

For the nonaphasics, the mean recall score for incidental learning was 10.66 with a standard deviation of 3.88. The mean recall score for goal-directed learning was 13.12 with a standard deviation of 2.55. The correlation between recall scores obtained with incidental and goal-directed instructions was .6135. A *t* test applied to the difference between the mean scores yielded a *t* of 3.93, which was found to be significant beyond the .001 level. These results indicated that nonaphasics performed significantly better with goal-directed than with incidental instructions.

For the aphasics, the mean recall score for incidental learning was 9.92 with a standard deviation of 3.17. The mean recall score for goal-directed learning was 9.06 with a standard deviation of 3.24. By inspection of the mean scores it is apparent that the aphasics did not earn a higher recall score with goal-directed instructions. Their mean recall score with goal-directed instructions is lower than the mean score earned with incidental instructions. These results support the hypothesis.

---

[3] The writer is indebted to the staffs of the neurology and psychology services of the Boston VA Hospital, the Bronx VA Hospital, and the Mental Hygiene Clinic of the VA New York Regional Office for their cooperation and assistance in making patients available for this study.

In order to determine whether the difference between recall scores obtained by the aphasics was significant, a *t* test was applied to the difference between the mean scores. The correlation between incidental and goal-directed scores was .7167. The *t* of 1.76 which resulted from this analysis, indicated that the difference was not significant.

Discussion

The results of the experiment show that goal-directed instructions, as compared with incidental instructions, produced higher recall scores in the nonaphasics but not in the aphasics. What is the explanation for the lack of improvement in the scores of the aphasics despite their knowledge of the task expected of them?

One possible explanation for this failure to profit from the goal-directed set may be the aphasic's relative inability to retain or maintain the set induced by the goal-directed instructions with sufficient consistency to perform more adequately. Such a notion is consistent with studies of brain functions in both animals and humans which report increased distractibility in the performance of brain-injured *S*s (4, 7, 9). The psychological processes responsible for this increased distractibility in the human *S*s, while difficult to specify, may be related to deficiencies in the use of implicit verbalizations or cue-producing responses. Implicit verbal cues tend to be employed for "rehearsal" of the instructions when the task becomes complicated. In such circumstances, one may quickly "run through the steps" of a demanded activity to make sure one understands the task. The goal-directed learning task in the present study was purposely made simple in order to minimize interferences which might result from this possible inability to maintain the set induced by instructions.

Another, and more far-reaching explanation for the aphasic's relative inability to profit from knowledge of the goal may be related to the aphasic's drive state when confronted with the need to solve a problem. Aphasics have a high degree of experience with failure. Not only are they handicapped with respect to the use of language, they also suffer from severe restrictions in the use of their limbs (many aphasics are paralyzed on one side of the body). Thus, they experience repeated failure in ordinary motor situations, in the process of communication and in social relationships which characteristically depend upon speech proficiency. It was thought, therefore, that as soon as aphasics are made aware of a new expectation of achievement, their history of experience with failure might reactivate anxieties which inhibit their deriving the same benefit from a goal-directed learning situation as nonaphasics can.

This notion of the inhibitory effects of anxiety is consistent with Goldstein's (4) theory of the organic patient's inability to cope with a problem as the determinant of his catastrophic reaction. Undoubtedly, the anticipation of this inability to cope, or the memory of previous failures, can serve to evoke or to perpetuate the feelings of anxiety which may have interfered with a more efficient performance of the aphasics in the goal-directed learning situation.

Either one of the two explanations offered can account for the data, but it is also possible that both of them are applicable. It is possible, for example, that anxiety due to the reactivation of previous experiences with failure may have made it more difficult for the aphasics to retain the set induced by the instructions.

In the present experiment, it was not possible to observe directly the drive state of the *S*s or even to subject it, indirectly, to some measurement procedure. It might be of interest, however, to note that one of the aphasics repeatedly requested that he be permitted to terminate the goal-directed learning situation, although he had made no such request on the previous day in the incidental learning situation.

Summary

The study investigated the relative effectiveness of goal-directed and incidental learning in aphasic patients. Observations about the performance of aphasics made by different workers were the source of the hypothesis which stated that in aphasic patients the normally more effective goal-directed learning

situation is not significantly more effective than the incidental learning situation. In order to test this hypothesis, two equivalent series of postage stamps were exposed to 25 predominantly expressive male aphasic patients between the ages of 19 and 51, once with instructions to remember the stamps and once with instructions to evaluate the stamps aesthetically. Retention of the material was tested and the scores obtained by the aphasics were compared with those obtained by a control group of 25 male nonaphasics. Results confirmed the hypothesis.

Received May 7, 1957.

References

1. Conrad, K. Das Problem der gestoerten Wortfindung in gestalttheoretischer Betrachtung. *Schweiz. Arch f. Neurol. Psychiat.*, 1949, **63**, 141–192.
2. Dollard, J., & Miller, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.
3. Freud, S. Psychopathology of everyday life. In A. A. Brill (Ed.), *Basic writings of Sigmund Freud.* New York: Modern Library, 1938. Pp. 35–150.
4. Goldstein, K. *The organism.* New York: American Book, 1939.
5. Goodenough, Florence L., & Anderson, J. E. *Experimental child study.* New York: Century, 1931.
6. Granich, L. *Aphasia.* New York: Grune & Stratton, 1947.
7. Malmo, R. B. Interference factors in delayed response in monkeys after removal of frontal lobes. *J. Neurophysiol.*, 1942, **5**, 295–308.
8. Sheehan, Vivian M. Techniques in the management of aphasics. *J. Speech Hear. Disorders*, 1948, **13**, 241–246.
9. Strauss, A. A., & Lehtinen, Laura E. *Psychopathology and education of the brain-injured child.* New York: Grune & Stratton, 1947.
10. Wepman, J. M. *Recovery from aphasia.* New York: Ronald, 1951.

Journal of Consulting Psychology
Vol. 22, No. 2, 1958

Wechsler-Bellevue Scatter as an Index of Schizophrenia

Arnold Trehub and Isidor W. Scherer
VA Hospital, Northampton, Mass.

Numerous studies have been reported which have used some measure of scatter on the Wechsler-Bellevue as a basis for clinical diagnosis. The recent review by Guertin, Frank, and Rabin (2) summarizes the yield to date as "inconclusive." The criticism has been offered that instances where good agreement has been found between psychiatric and W-B diagnoses may be "due to the inspectional techniques of the clinician rather than the numerical scatter values" (4, p. 229). Furthermore, it is held that while W-B scatter may successfully differentiate *groups* in terms of a statistically significant difference, it has not been demonstrated to be effective in the diagnosis of *individuals*.

The present study dealt with the individual discrimination of schizophrenics from neurotics and character disorders on the basis of a scatter index computed directly from W-B subtest scores.

Procedure

The sample consisted of a random selection of male patients in a Veterans Administration neuropsychiatric hospital who held a psychiatric diagnosis of schizophrenia, neurosis, or character disorder, who had been referred for psychological testing and who had been administered a complete (11 subtests) W-B Form I. The scatter index used involved sheer variability of subtest performance without reference to any particular patterns of achievement. Weighted scores of all eleven subtests on the W-B were averaged for each case, and the differences of all subtests from the mean were summed disregarding sign. The sum obtained was the individual's scatter score. An earlier pilot investigation showed that when all patients in the sample who had a scatter score of 19 or greater were labeled schizophrenic and all below 19 were labeled neurotic or character disorder, we were correct in approximately 70% of the cases against the criterion of psychiatric diagnosis. Accordingly, in the present study a sequential analysis plan was set up (1) in which the hypothesis of 60% correct diagnoses (H_o:$P_o = .60$) was tested against the alternate hypothesis of 70% correct diagnoses (H_1:$P_1 = .70$). Risks α and β were both set at .05. As each case was drawn, the diagnostic decision was based upon the following rules:

Scatter index $\geqq 19$ = schizophrenia.
Scatter index < 19 = neurosis or character disorder.

Each diagnosis made on the basis of the scatter index was matched against psychiatric diagnosis and plotted as correct or incorrect in the sequential analysis channel.

Results and Discussion

When the 269th case was plotted, the sequential analysis plan yielded the decision to reject the hypothesis of 60% correct diagnoses in favor of the hypothesis of 70% correct diagnoses ($\alpha = .05$, $\beta = .05$).

Computation of mean age at the time of testing shows the mean age in years to be 31.8 for the schizophrenics, and 32.9 for the neurotics and character disorders. Mean IQ is 103.4 for the schizophrenics and 111.0 for the neurotics and character disorders.

Since the patients used in this study constituted a random sample of schizophrenics, neurotics, and character disorders who were referred for psychological testing, the diagnostic composition of the total sample may be considered to approximate the actual base

Table 1

Scatter Index Cutting Scores and Associated Diagnostic Consequences

Scatter Index Cutting Score	Schizophrenics			Neur. and Charac. Disorders							
	Frequency	Cumulative Frequency	p_1	Frequency	Cumulative Frequency	p_2	P_{p_1}	Q_{p_2}	H_T	H_P	R_P
$\geqq 28$	20	20	.1205	1	1	.0097	7.4	.4	45.3	94.9	7.8
27	6	26	.1566	1	2	.0194	9.6	.7	47.2	93.2	10.3
26	10	36	.2169	3	5	.0485	13.4	1.9	49.8	87.6	15.3
25	8	44	.2651	3	8	.0777	16.4	3.0	51.7	84.5	19.4
24	14	58	.3494	3	11	.1068	21.6	4.1	55.8	84.0	25.7
23	11	69	.4157	5	16	.1553	25.6	5.9	58.0	81.3	31.5
22	10	79	.4759	3	19	.1845	29.4	7.1	60.6	80.5	36.5
21	11	90	.5422	3	22	.2136	33.4	8.2	63.5	80.3	41.6
20	16	106	.6386	7	29	.2816	39.4	10.8	66.9	78.5	50.2
19	9	115	.6928	9	38	.3689	42.7	14.1	66.9	75.2	56.8
18	8	123	.7410	4	42	.4078	45.7	15.6	68.4	74.6	61.3
17	13	136	.8193	6	48	.4660	50.6	17.8	71.1	74.0	68.4
16	7	143	.8614	4	52	.5048	53.1	19.3	72.1	73.3	72.4
15	4	147	.8855	12	64	.6214	54.6	23.8	69.1	69.6	78.4
14	3	150	.9036	11	75	.7282	55.8	27.9	66.2	66.7	83.7
13	4	154	.9277	15	90	.8738	57.2	33.5	62.0	63.1	90.7
12	7	161	.9699	7	97	.9417	59.8	36.1	62.0	62.4	95.9
11	2	163	.9819	3	100	.9709	60.6	37.2	61.7	62.0	97.8
$\leqq 10$	3	166	1.0000	3	103	1.0000	61.7	38.3	61.7	61.7	100.0

Note.—Base rate for schizophrenics (P) = 61.7%. Base rate for neurotics and character disorders (O) = 38.3%.

rates of such patients at our hospital. We found that schizophrenics composed 61.7% of our sample, while neurotics and character disorders composed 38.3%. This breakdown, of course, is based upon a hospital subpopulation from which other diagnostic classifications such as brain damage, psychotic depression, etc. are excluded.

It has been shown that with the method of computing scatter used in this study, and with the adoption of a scatter index of 19 as a cutting score, schizophrenics could be differentiated from neurotics and character disorders in more than 60% of the cases. In terms of the sequential analysis model, operating at a level of 70% correct diagnoses is significantly more likely than operating at a level of 60% correct diagnoses. But, having established this, we are still concerned with a number of other questions about the diagnostic decisions that might be made on the basis of the W-B scatter index. Table 1 presents an analysis of diagnostic consequences in the sample of patients for a number of different scatter index cutting scores. It is based upon procedures outlined by Meehl and Rosen (3). Following are definitions of terms used in the table:

p_1 = Proportion of schizophrenics correctly identified by the test ("valid positive" rate).
p_2 = Proportion of neurotics and character disorders misidentified by the test as schizophrenic ("false positive" rate).
P_{p_1} = Percentage of total sample correctly identified as schizophrenic.
Q_{p_2} = Percentage of total sample misidentified as schizophrenic.
H_T = Percentage of total correct diagnoses.
H_P = Percentage of correct diagnoses if a diagnosis is made only for patients scoring at or above cutting score.
R_P = Percentage of patients in total sample who receive a diagnosis if a diagnosis is made only for patients scoring at or above cutting score.

If nothing but base rates were available to us for making a diagnostic decision, we would call all patients schizophrenic and we would be correct in 61.7% of the cases. In this regard, one might ponder how much an informal appreciation of patient base rates might be contributing to our diagnostic acumen.

It can be seen from Table 1 that when our original cutting score of 19 is used, we are correct in 66.9% of the cases (H_T). We have thus shown an improvement of 5.2% over base rate diagnosis. However, when we use a cutting score of 16, our percentage of total correct diagnoses rises to 72.1%, an improvement of 10.4% over base rate diagnosis. If we choose to make a diagnostic statement only for those patients identified as schizophrenic by our index, it can be seen that with a cutting score of 19 we make 75.2% correct diagnoses (H_P). In this case, however, we are able to offer a diagnosis for only 56.8% of the patients in our total sample (R_P). Within the limits of our procedure, as we increase our accuracy in making positive diagnoses, we pay the penalty of reducing the percentage of patients about whom we can make diagnostic statements. Numerous other contingencies may be observed from Table 1. The reliability and generality of our "valid positive" and "false positive" rates can only be determined by cross-validation in a variety of clinical settings. Their associated diagnostic contingencies will vary as actual clinical base rates P and Q change.

Summary

A Wechsler-Bellevue scatter index based upon the sum of subtest deviations from the mean subtest score was computed for a sample of schizophrenics, neurotics, and character disorders. A cutting score of 19 was used as a basis for discriminating schizophrenics from neurotics and character disorders. The accuracy of discrimination was judged against the criterion of psychiatric diagnosis. The hypothesis of 60% correct diagnoses was rejected in favor of the alternate hypothesis of 70% correct diagnoses in a sequential analysis test. Data were presented showing other diagnostic consequences in our sample of patients arising from the application of various scatter index cutting scores.

Received April 25, 1957.

References

1. Fiske, D. W., & Jones, L. V. Sequential analysis in psychological research. *Psychol. Bull.*, 1954, **51**, 264–275.
2. Guertin, W. H., Frank, G. H., & Rabin, A. I. Research with the Wechsler-Bellevue intelligence scale: 1950–1955. *Psychol. Bull.*, 1956, **53**, 235–257.
3. Meehl, P. E., & Rosen, A. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychol. Bull.*, 1955, **52**, 194–216.
4. Rabin, A. I., & Guertin, W. H. Research with the Wechsler-Bellevue test: 1945–1950. *Psychol. Bull.*, 1951, **48**, 211–248.

Journal of Consulting Psychology
Vol. 22, No. 2, 1958

Validity of the Hewson Ratios: Investigation of a Fundamental Methodological Consideration[1, 2]

Walter F. McKeever and Alvin I. Gerstein
University of Rochester

Hewson (2) has described a method for diagnosing brain pathology from ratios computed between Wechsler-Bellevue I weighted subtest scores. The effects of age and IQ variables on these ratios have never been assessed, and little validation data are available on the method. The present study aims to help fill these gaps.

*S*s were 50 schizophrenics and 52 organics. All were between 18–55 years old, had had no electroshock, had full W-B I tests, and had IQs of at least 80. Adequacy of diagnosis was partially assured by the requirement that each patient have at least two congruent, independent diagnoses, and never have been diagnosed differently. The total groups were matched on age and IQ. These groups were also divided into those with IQs above and below 100, so that 30 schizophrenics and 30 organics made up the 100+ groups, the remainders constituting the 99− groups. There were no differences within or between groups on age, and none on IQ across diagnostic groups.

Over-all accuracy of Hewson diagnoses was 54.9%, or essentially chance. Although the 100+ groups were significantly differentiated in the correct direction, fully a third of the 100+ schizophrenics were misclassified as organics. In the 99− groups, actual organics were correctly identified no better than chance, while 75% of the schizophrenics were misclassified as organics, a finding approaching significance ($p = .10$). Patients who performed organically were significantly older and less intelligent as a group than were those who did not ($p = .01$). Point biserial correlations between organic performance and age and IQ, respectively, were .45 and − .34, both significant beyond the 1% level. The r between age and IQ was .09, nonsignificant.

These results are generally inconsistent with the more positive ones reported by Gutman (1). Inspection of Gutman's data suggests the possibility that her organics may have been significantly older and less intelligent than the controls, a situation favoring artifactually positive results.

The following conclusions seem justified:

1. Organic performance on the Hewson ratios varies systematically with age and IQ variables. Nearly a third of the variance of organic performance is associated with variance in age and IQ.

2. The method does not differentiate the groups studied beyond a chance level; among patients with IQs below 100 it makes at least as many errors as it does correct diagnoses.

3. The need for control of logically relevant variables seems re-emphasized by the findings.

Brief Report.
Received December 2, 1957.

References

1. Gutman, Brigette. The application of the Wechsler-Bellevue Scale in the diagnosis of organic brain disorders. *J. clin. Psychol.*, 1950, **6**, 195–198.
2. Hewson, Louise R. The Wechsler-Bellevue Scale and the Substitution Test as aids in neuropsychiatric diagnosis. *J. nerv. ment. Dis.*, 1949, **109**, 158–183, 246–266.

---

[1] An extended report of this study may be obtained without charge from Walter F. McKeever, Department of Psychology, University of Rochester, Rochester, N. Y., or for a fee from the American Documentation Institute. Order Document No. ——, remitting $—— for microfilm or $—— for photocopies.

[2] The aid of B. F. McNeal and the staff of Canandaigua VA Hospital, Canandaigua, N. Y., is gratefully acknowledged. The study was done during the writers' internships at Canandaigua.

Journal of Consulting Psychology
Vol. 22, No. 2, 1958

Brain Injury and Intellectual Performance

Alan O. Ross[1]
Clifford Beers Guidance Clinic, New Haven, Conn.

The question whether intelligence deteriorates after brain injury has been investigated in a number of studies whose results are contradictory (4). In connection with an investigation dealing with the effect of brain injury on tactual perception, this writer (3) gathered data relevant to intelligence test performance which, though not conclusive, indicate that brain injury may have a deleterious effect on the ability measured by such tests.

Subjects (*S*s) were 20 soldier patients with a mean age of 22.8 years who had undergone brain surgery within 12 months before the date of the study. Only those cases were included where surgery had involved at least an incision of the dura. Except for one tumor and two accident cases, all traumata had been due to gunshot or shellfragment injuries located in the frontal, temporal, or parietal regions. For each *S*, the score he had made on the Army General Classification Test (5), taken some time before he sustained his injury, was obtained from the soldier's records, furnishing a measure of his premorbid intelligence. At the time of the study, each *S* was given the CVS Individual Intelligence Scale (1) which provided a measure of his postinjury functioning. The CVS consists of the comprehension and similarities items of the Wechsler-Bellevue and a vocabulary scale based on the Stanford-Binet word list. The CVS correlates with a classification test simliar to the AGCT on the order of .80 (1), and in an analysis conducted by this writer, using 37 normal *S*s, the AGCT and CVS were found to have an r of .75 ($t = 6.61$, $p < .001$).

By transforming the AGCT and CVS scores into deciles, it is possible to test for the significance of the difference between the patients' preinjury and postinjury scores, using the formula for correlated means presented by McNemar (2). The postinjury scores are significantly lower than the preinjury scores with a t of 3.50 ($p < .01$). This relationship can also be demonstrated by a matched group technique where the more highly discriminating CVS standard scores and original AGCT scores can be utilized, thus eliminating the transformation into deciles. Each patient was matched individually with a normal, drawn from the service personnel of a major army hospital, controlling for AGCT score, age, and years of education. Each of the normals was given the CVS as part of the study. Table 1 shows the result of the AGCT matching and the comparison of CVS scores for patients and normals. When patients are matched with normals on the basis of the patients' preinjury scores, the patients have significantly lower scores when tested within 12 months after brain surgery.

A converse comparison is possible by matching the patients with different normals using the criterion of patients' postinjury (CVS) scores, again holding age and years of education constant. An analysis of the differences in AGCT scores for the matched pairs shows that the patients' (preinjury) scores are significantly higher with a t of 3.23 ($p < .01$).

These results, though statistically significant, provide only presumptive indication that intelligence test performance deteriorates as the result of certain forms of brain injury. A more direct test of this question would have been to readminister the AGCT to all patients postoperatively, but this was unfortunately not possible due to the procedure dictated by

---

[1] At the time of this study the writer was a 1st Lt., MSC, US Army, stationed at Walter Reed Army Hospital.

Table 1

Intelligence Test Scores of Patients and Normals ($N = 20$ pairs)

	Mean		*t* for Mean	
Test	*Normals*	*Patients*	*Difference*	*p*
AGCT	59.0	59.6	.62	$>.50$
CVS	31.4	28.4	2.50	$<.05$

the primary study (3) to which these findings are incidental.

Received April 29, 1957.

References

1. Hunt, W. A., & French, Elizabeth G. The CVS abbreviated individual intelligence scale. *J. consult. Psychol.,* 1952, **16**, 181–186.
2. McNemar, Q. *Psychological statistics.* New York: Wiley, 1949.
3. Ross, A. O. Tactual perception of form by the brain-injured. *J. abnorm. soc. Psychol.,* 1954, **49**, 566–572.
4. Ross, A. O. Integration as a basic cerebral function. *Psychol. Rep.,* 1955, **1**, 179–202. (Monogr. Suppl. No. 2.)
5. U. S. War Department, Adjutant General's Office, Classification and Replacement Branch, Pers. Res. Sec. The army general classification test. *Psychol. Bull.,* 1945, **42**, 760–768.

Journal of Consulting Psychology
Vol. 22, No. 2, 1958

The Clinical Usefulness of the Archimedes Spiral in the Diagnosis of Organic Brain Damage[1, 2]

Lewis R. Goldberg
University of Michigan

and Philip A. Smith
Ann Arbor VA Hospital

In reporting their promising results using the spiral aftereffect to differentiate groups of patients with cortical involvement from nonorganic control groups, Price and Deabler concluded that ". . . the results are such as to justify present clinical use of this technique in diagnosis of organicity" (4, p. 302). They found almost no overlap in the performance of their groups: fewer than 8% of normals or functional psychiatric patients were misclassified as organic and only 2% of their organic group was incorrectly classified. While they reported no data on such variables as age, which might have contributed to their results, the description of the syndromes included in the organic group suggest that this group was considerably older than the normals or psychiatric patients. However, later studies by Gallese (2) and Page *et al.* (3) showed no significant age-score relationship, and offered some additional evidence for Price and Deabler's original assumption. Gallese, however, noted that the spiral test was relatively insensitive with certain types of organic patients, notably those with convulsive disorders or brain syndromes associated with alcoholism. The performance of twelve lobotomized schizophrenics also was indistinguishable from that of nonlobotomized schizophrenics and normals. Page *et al.* supported Price and Deabler's theoretical assumptions but their study provided ". . . less evidence that the aftereffect may serve as an effective diagnostic device" (4, p. 91). These investigators also noted unimpaired performance on the part of some lobotomized schizophrenics. They did not report separate data for patients with convulsive disorders. A study by Standlee (5) of psychiatric patients tested before and after electric shock treatment (which generally produces some temporary organic dysfunction) was inconclusive because the effect of prior experience with the spiral was not controlled.

During the course of their clinical practice at the Veterans Administration Hospital, Ann Arbor, Michigan, the present writers utilized the Archimedes spiral as a part of their test battery in examining patients referred for psychological evaluation. The present paper reports on findings over a period of one year.

Method

Apparatus. The instrument used in this study was a 78 rpm phonograph turntable, vertically mounted on an 8 × 14 inch black box which housed the motor. One clockwise and one counterclockwise Archimedes (or Plateau) spirals, each $2\frac{1}{2}$ circuits (920 degrees) about the center, were painted in black on 10-inch white cardboard discs, which could be affixed to the turntable. This instrument differed slightly from that of

---

[1] The authors wish to thank E. Lowell Kelly of the University of Michigan for his critical reading of this manuscript, and express gratitude to the Psychiatric Research Committee of the Ann Arbor VA Hospital for encouragement and assistance in this study.

[2] The statements and conclusions of the authors do not necessarily reflect the opinions or policy of the Veterans Administration.

Table 1

Diagnostic Classification of Subjects

Subject group	Diagnosis	*N*
Normal	No pathology	30
Psychiatric	Schizophrenia	4
	Psychoneurosis	8
	Psychophysiologic reaction	5
Post-EST	Schizophrenia	10
	Psychoneurosis	1
Organic	Cerebral thrombosis	7
	Arteriosclerosis	5
	Convulsive disorder	5
	Cerebral neoplasm	3
	Encephalomyelitis	2
	CNS lues	1
	Paralysis agitans	1

previous investigators both in speed of rotation and diameter of the spiral. Others (1, 2, 3, 4) have reported using spirals varying from 3¾ inches to 8 inches in diameter and at speeds varying from 78 to 100 rpm.

Procedure. The following procedure was standard for all subjects (*S*s): *S* was seated approximately 8 feet from the instrument in an office with good illumination. The spirals were presented in ABBA order, with the interval between successive trials varying with the duration of reported after-image. Each trial was for 30-seconds' duration. Instructions given to *S*s were essentially the same as those used by Price and Deabler: "This is a special eye test. Look at the center (pointing) and don't take your eyes away until I tell you to." Approximately ten seconds after rotation was begun, *S* was asked, "What does the line appear to be doing?" After the spiral had been braked to a stop, the last instruction was repeated and *S*'s response was recorded.

Subjects. The population from which *S*s were selected included both staff and patients from the psychiatric and neurological wards of the Ann Arbor VA Hospital (Table 1). Patients classified as organics were those with mild to grossly limiting cortical dysfunction, confirmed by independent neurological diagnosis. Patients with questionable or marginal organic dysfunction were not included in the present analysis, except for a group of eleven psychiatric patients who were seen one to seven days following a course of four or more electro-shock treatments (EST). These were presumed to show a possible temporary organic deficit due to the effects of EST. The remaining psychiatric patients were untreated by EST and were seen during the acute fulminating stage of their illness, shortly after admission to the hospital. The normal control group included ward physicians, psychiatrists, psychologists, nurses, secretaries, and students employed at the hospital.

Scoring. Responses from all *S*s were recorded verbatim and scored according to the criteria of Price and Deabler and also those proposed by Gallese. The latter method, which gave equal credit for a description of the aftereffect either as a change in apparent size or direction of movement, was judged the most applicable for the present analysis, since a large proportion of responses was phrased as some combination of these terms. Working independently from a sample of 38 verbatim protocols, the authors agreed perfectly in this manner of scoring for 35 *S*s. Since four partial scores contributed to the total score for each *S*, this represented a total of 152 sets of independent judgments for which the scorers agreed 146 times (96%).

Results

Table 2 summarizes the findings for all groups. The normal *S*s were superior to all others, attaining, without exception, perfect scores. Because of the absence of within-

Table 2

Spiral Aftereffect Scores

Subject group	Mean age	Mean score	Mean score adjusted for age
Normal	29.3	4.00	—
Psychiatric	35.7	2.94	2.72
Post-EST	38.5	2.55	2.45
Organic	45.0	2.17	2.37

Table 3

Distribution of Aftereffect Scores
(after Price and Deabler)

	Score				
Subject group	0	1	2	3	4
Normal	0%	0%	0%	0%	100%
Psychiatric	11%	6%	18%	6%	59%
Post-EST	18%	0%	27%	18%	37%
Organic	33%	4%	13%	13%	37%

group variability for the normal *S*s, usual parametric tests of significance were inappropriate. Accordingly, a chi-square value was obtained for the difference between normals and the combined patient groups. When this was found to be significantly large ($\chi^2 = 23.50$, $p < .001$), the normals were excluded from subsequent statistical tests. An analysis of covariance for spiral scores of the remaining groups, adjusted for age, was not statistically significant ($F < 1$). Examination of the adjusted mean scores, presented in Table 2, shows that the organics nevertheless maintained the low order position among the groups.

The Pearson product-moment correlation between age and spiral score was $- .39$. For 32 patients for whom IQ measures were available, the correlation between intelligence and spiral score was small and not significant ($r = .11$).

Because these findings seemed so incompatible with the results of previous studies, the data were arranged according to the procedures followed by earlier investigators and new comparisons made. Table 3 shows the distribution of scores following Price and Deabler's original presentation. The patients were then split into those whose scores were 2 or under and those whose scores exceeded 2, after Gallese. Table 4 summarizes this analysis and shows the percentage of patients who were correctly classified as "organic" or "normal" by this method. Table 4 also shows the results of this breakdown for patients classed as *convulsive disorder* or *"other types"* of organic dysfunction. The present figures contrast sharply with those reported by Price and Deabler and Gallese.

Discussion

These findings cast considerable doubt as to the utility of the spiral aftereffect for differential diagnosis of organicity. Although Price and Deabler's results appeared so promising, later work (2, 3) made their findings look more conditional. The present findings add an even more compelling note of caution to the use of the instrument as a method of diagnosis. This seems especially important since all the published studies used clear-cut cases of cortical dysfunction, while in clinical practice, neurological involvement is often marginal or partial and thus the diagnostic problem might be expected to be even more difficult.

The organics used in the present study were all drawn from a group of recent admissions to a general medical and surgical hospital, and one might hypothesize that such patients tend more likely to show acute rather than chronic deteriorative effects of cortical destruction. Patients exhibiting chronic symptoms are more typically found in the VA neuropsychiatric hospitals and state institutions where the earlier studies were conducted. This might be one reason why the present data differs from previous findings. For the same reason, however, the present psychiatric groups did not include chronic, deteriorated schizophrenics, and it is difficult to understand how so large a proportion of these relatively intact patients failed to attain perfect scores. Our instructions and scoring procedures were identical to those of earlier investigations although some differences may have arisen from the manner in which

Table 4

Classification of Subjects According to
Gallese Criteria

Subject group	Organic score (2 or under)	Normal score (3 or over)	Correctly classified
Normal	0	30	100%
Psychiatric	6	11	65%
Post-EST	5	6	—
Organic	12	12	50%
Convulsive disorder	2	3	40%
Other types	10	9	53%

inquiries were conducted. Gallese noted that many schizophrenics would have been scored "organic" had not a detailed, direct inquiry been performed, and Page *et al.* reported a need for an objective nonverbal response measure for use with the spiral. Perhaps a multiple choice response situation such as that offered by Freeman and Josey (1) would insure more uniform results.

Many patients, both organic and psychiatric, suffered from disturbances in their ability to communicate effectively. Thus, although the spiral as originally presented appeared to be a neat, objective evaluative technique, in actual practice a large degree of clinical acumen may be necessary to determine an *S*'s exact perception. The less contact the patient has with reality (whether functional or organic in origin) the less is the likelihood that correct inferences regarding his perception can be made. Caution must be observed in interpreting the remarks of negativistic patients, as well as overly defensive persons who wish to deny any experience which they perceive as a malfunction. Some patients expressed fear that reporting the aftereffect was tantamount to admitting a delusion, and one was concerned lest he be hypnotized unknowingly. Other patients, despite repeated instructions, responded to the spiral as a kind of projective technique, reporting such phenomena as "it looks like a coiled snake, ready to strike," and "it is a ram's horn, all curled up." The present authors do not imply that these cautions are in any way unique to the spiral technique but, rather, that this test shares with many other diagnostic instruments the difficulties encountered in evaluating verbal report.

It would appear also that variables such as the rate of rotation and size of the spiral, level of illumination, and other test conditions might well have some influence on results. For example, it was noted in preliminary work that the perception of movement toward or away from *S* was in part a function of the distance between the exact center of the turntable and the placement of the spiral upon it. When the spiral was positioned so that it turned about its *exact* center, then much less perception of changing distance occurred. Whether such variables have a differential effect on different populations is unknown.

The one variable in this study which did seem related to ability to perceive the effect was *age*. Moreover, Freeman and Josey found marked differences in ability to see the aftereffect in a population of psychotic patients, and although they related this deficiency to memory impairment, their data would indicate that age may have been a factor. In both the Price and Deabler and Gallese studies, the organic groups appeared substantially older than their nonorganic controls.

There is a discrepancy between the present results and those of Standlee in regard to post-EST patients. Standlee reported that 23 out of 25 patients perceived the aftereffect on a retest 8 hours after one EST. In the present post-EST groups, scores ranged from 0 to 4 and were little correlated with clinical impression of "organicity." As has been suggested elsewhere (4), practice effects may have contaminated Standlee's results; moreover, the discrepancy between the two studies was probably even more heightened by the fact that all Standlee's *S*s had considerably fewer shock treatments than the present ones. On the other hand, in the present study there was no apparent correlation between aftereffect score and number of EST or length of time since the last treatment.

One of the paramount problems in all studies utilizing the spiral has been the difficulty in establishing a satisfactory set of criteria of organic brain damage. In most studies, an implicit assumption is made that localization, chronicity, and degree of neurological destruction are irrelevant variables, as well as such factors as the optical efficiency of the *S* and the pharmacological effect of the drugs he has been administered. The fact that such a viewpoint is basically unsound has already been suggested (6). To date most studies, including the present one, confound these crucial variables and thus they cannot be effectively compared.

Summary

This study is an evaluation of the Archimedes spiral as a clinical technique for the diagnosis of organic brain damage. Over a pe-

riod of a year's work with the instrument the authors noted that normal *S*s report perceiving the aftereffect of both expanding and contracting spirals without a single instance of failure. Psychiatric, post-EST, and organic patients—in respective order—performed with decreasing efficiency on the same task. When the scores for the groups were adjusted for age, however, the differences between the latter three groups became statistically indistinguishable. The correlation between age and spiral score was − .39.

In spite of early enthusiasm for the spiral, the present results warn against its indiscriminate use for differential diagnosis.

Received June 1, 1957.

References

1. Freeman, E., & Josey, W. E. Quantitative visual index to memory impairment. *A.M.A. Arch. Neurol. Psychiat.*, 1949, **62**, 794–797.
2. Gallese, A. J. Spiral aftereffect as a test of organic brain damage. *J. clin. Psychol.*, 1956, **12**, 254–258.
3. Page, H. A. *et al.* Another application of the spiral aftereffect in the determination of brain damage. *J. consult. Psychol.*, 1957, **21**, 89–91.
4. Price, A. D., & Deabler, H. L. Diagnosis of organicity by means of spiral aftereffect. *J. consult. Psychol.*, 1955, **19**, 299–302.
5. Standlee, L. S. The Archimedes negative aftereffect as an indication of memory impairment. *J. consult. Psychol.*, 1953, **17**, 317.
6. Yates, A. J. The validity of some psychological tests of brain damage. *Psychol. Bull.*, 1954, **51**, 359–380.

Journal of Consulting Psychology
Vol. 22, No. 2, 1958

Identification of Item Factor Patterns Within the Manifest Anxiety Scale[1]

A. W. Bendig
University of Pittsburgh

O'Connor, Lorr, and Stafford (2) have reported a factor analytic study of the item content of Taylor's Manifest Anxiety Scale. Tetrachoric correlations were computed among 42 MAS items (omitting eight items with marginal frequencies of less than 10 per cent) for a sample of 220 undergraduate *S*s. Five oblique factors were extracted and described in terms of the apparent content of items highly loaded for each factor. No external validation of the factors was attempted.

It has been shown (1) that MAS scores correlate highly ($r = .76$) with scores from the Neuroticism (N) scale contained in Eysenck's Maudsley Personality Inventory (MPI) and are less highly related ($r = -.34$) to the MPI Extraversion (E) scale. This suggests that some of the MAS factor patterns found by O'Connor, Lorr, and Stafford may be identified as "extraversion" or "neuroticism" factors within the multidimensional MAS items. Since the factor analytic study used mixed male and female *S*s it is possible that one or more of the factors obtained might be a "sex" factor.

Both the MAS and MPI inventories were administered to 145 *S*s (100 men and 45 women) enrolled in introductory psychology. Each of the 42 MAS items was correlated (biserial-phi) with the sex criterion and these coefficients were correlated (product-moment) with the item factor loadings reported by O'Connor, Lorr, and Stafford. None of these five correlations (one for each factor) were significant.

The *S*s were dichotomized at the medians of the distributions of their E and N scores and each of the 42 MAS items was correlated (tetrachoric) with total scores on these two MPI scales. The E criterion item validities were then correlated with each of the five sets of item factor loadings and a similar procedure was followed with the N criterion item validities. The item validities for the E scale criterion were significantly related (.01 level) to the factor loadings on Factor D ($r = -.51$), but were not significantly correlated with the other four factors ($r = .06$, .01, .05, and $-.04$). The N scale item validities were significantly correlated with factor loadings on Factors A, B, and D ($r = .38$, $-.49$, and .35), but were not related to Factors C and E ($r = .23$ and .25). The E and N correlations with Factor A ($r = .06$ and .38) were significantly different at the .01 level ($F = 7.94$) as were the E and N correlations with Factor B ($r = .01$ and $-.49$, $F = 11.12$). The E and N correlations with Factor D ($r = -.51$ and .35) were not significantly different ($F = 1.06$) because of the negative correlation ($r = -.42$) between the E and N criteria validities of the 42 MAS items.

Factors A and B found by O'Connor, Lorr, and Stafford (2) appear to be similar to the "neuroticism" factor identified by Eysenck, while Factor D seems to be a combination of Eysenck's "neuroticism" and "introversion" factors.

Brief Report.
Received November 12, 1957.

References

1. Bendig, A. W. Extraversion, neuroticism, and manifest anxiety. *J. consult. Psychol.*, 1957, 21, 398.
2. O'Connor, J. P., Lorr, M., & Stafford, J. W. Some patterns of manifest anxiety. *J. clin. Psychol.*, 1956, 12, 160–163.

---

[1] An extended report of this study may be obtained without charge from A. W. Bendig, Dept. of Psychology, University of Pittsburgh, Pittsburgh 13, Pa., or for a fee from the American Documentation Institute. Order Document No. 5466, remitting $1.25 for microfilm or $1.25 for photocopies.

Journal of Consulting Psychology
Vol. 22, No. 3, 1958

Intelligence in Childhood Schizophrenics, Other Emotionally Disturbed Children, and Their Mothers

Anthony Davids

Brown University and Emma Pendleton Bradley Hospital

Much of the research devoted to assessment of intellectual functioning in children and adults has been generated by the nature-nurture or heredity-environment controversy (1). In this regard, several previous studies have been concerned with intellectual similarities between (*a*) normal children and their parents, (*b*) normal siblings, (*c*) identical twins and fraternal twins, and (*d*) foster children and their foster parents. The effects on children's IQ of impoverished home environments and of institutionalization have been studied. And the effects of enriched environments and special schooling have also been related to intellectual functioning in children. Since excellent surveys of these numerous studies, with exhaustive bibliographies, are available elsewhere (1, 5, 11), they will not be given detailed consideration in this report.

In spite of the wealth of previous investigation in this topical area, there is still very little empirical evidence bearing on the problem of relations between intellectual functioning in severely disturbed children and in their mothers. In the attempt to gain further understanding in this regard, the present study seeks answers to the following four questions. Is there a significant correlation between the IQs obtained by children who are severely disturbed emotionally and the IQs obtained by their mothers? Is there a significant difference between the IQs of the mothers and the IQs of their disturbed children? Is there a significant difference between the IQs of mothers of schizophrenic children and the IQs of mothers of children suffering from other forms of emotional disturbance? Is there a significant difference between the IQs of schizophrenic children and the IQs of other types of emotionally disturbed children? The findings obtained in answer to these questions should be of interest both to investigators who are concerned with theoretical aspects of the problem of psychopathology and intellectual functioning, and to clinicians who are faced with the practical problem of building programs for assessing children and parents in the many new child treatment centers that are being established.

Method

The findings to be presented in this report are the result of a five-year program of psychological assessment of hospitalized emotionally disturbed children and their mothers.[1] The patients were 70 boys and girls, between the ages of 5 and 12, who were undergoing residential psychiatric treatment at the Emma Pendleton Bradley Home. Of this sample, 12 children were diagnosed as schizophrenic, and 58 were diagnosed as either severe behavior disorder or neurotic. Also serving as subjects in this study were the mothers of the children. These 70 women, between the ages of 26 and 46, were all living in the community. At the time of the child's admission to the institution, the mother was examined with a comprehensive battery of psychological tests, including the Wechsler Bellevue Intelligence Scale (short form), Rorschach test, TAT, and figure drawings. The children were also examined by experienced clinical psychologists who included the complete Wechsler Intelligence Scale for Children, along with projective tests and visual-motor tests, in their battery of techniques. The present report is concerned solely with the results of intelligence testing; the results of the mother-and-child personality assessment must await further work. Thus, the data for this study consist of the mothers' IQs based on their performance on the comprehension, vocabulary, picture completion, and block design subtests of the adult W-B, and the children's IQs based on their responses to the 10 subtests of the WISC.

---

[1] The assessment program was initiated by M. W. Laufer, E. Z. Rubin, and S. Koret. The author wishes to gratefully acknowledge the many hours of testing conducted by Rubin, M. Talmadge, Pearl Gore, and several psychology interns who have served at the Bradley Home. The conception of the present study, experimental design, statistical analyses of data, and this report are the contributions of the present author. The writing of this report was facilitated by Research Grants M-1722 and M-2087 from the National Institute of Mental Health, United States Public Health Service.

Results and Discussion

The results show that there is a significant positive correlation ($r = .34$, $p = .01$) between the IQs of the 58 nonschizophrenic disturbed children and their mothers. In the schizophrenic group, however, there is no association ($r = .02$) between the indices of intelligence in the mothers and children. It should be pointed out that this correlation is based on the scores of only 8 schizophrenic children, as 4 of the group of 12 schizophrenics were untestable. They would neither attend to the test material nor attempt to cope with it, and in these cases only fragmentary, if any, test data were obtained, making any estimate of IQ practically impossible and even more unreliable than the estimates for the other schizophrenic children.

The coefficient obtained with the other types of disturbed children and their parents is lower than the coefficients in the vicinity of + .50 that have been reported (2, 3, 7, 8, 10) in studies of normal children and their parents, but even this degree of association is quite striking in view of the host of factors that would militate against finding a significant relation in our study. To mention only a few of these sources of error—the mothers' IQs were based on a shortened form of the intelligence test, the mothers were examined under conditions of stress (time of admission of child to hospital), and it is a well-known fact that emotionally disturbed children rarely are able to function at their optimal level on intelligence tests. Thus, in spite of these negative factors, a significant correlation was obtained, and one can only wonder what magnitude of relation might be found if these estimates of intelligence were not influenced by the intrusion of severe psychopathology.

Turning now to the differences between the performances of the mothers and children, presented in Table 1, a striking difference of 38 IQ points is found between the schizophrenics and their mothers, and a difference of 15 IQ points between the other disturbed children and their mothers. In both compari-

Table 1

Comparison of IQs of Mothers and of Their Emotionally Disturbed Children

Group	*N*	Mean		Variance	
Mothers of nonschizophrenics	58	108.1		175.8	
Mothers of schizophrenics	12	122.1		88.1	
Nonschizophrenic disturbed children	58	92.7		177.1	
Schizophrenic children	8	84.0		302.6	
Groups compared			*t*		*F*
Mothers of nonschizophrenics vs. mothers of schizophrenics			3.86**		2.00
Mothers of nonschizophrenics vs. their children			6.70**		1.01
Mothers of schizophrenics vs. their children			5.68**		3.44*
Nonschizophrenic disturbed children vs. schizophrenic children			1.65*		1.71

* Significant at .05 level.
** Significant at .001 level.

sons, these differences are significant well beyond the .001 level. Thus, it is evident that in the intellectual sphere the mothers fare much better than do their suffering offspring. It is clear that the burden of emotional conflict exacts a heavy intellectual toll in the young.

The two groups of mothers were compared. Even though the mean IQ of 108 obtained by the mothers of nonschizophrenic children is above the average for the general population, the mean IQ of 122 for the mothers of the schizophrenic children is significantly higher. This empirical finding is in keeping with clinical observations reported by several specialists in the field of childhood schizophrenia. Kanner, in describing the "indisputably high level of intelligence" of parents of cases of infantile autism, reports that "All but 5 of the mothers of the 55 children have attended college. All but one have been active vocationally before, and some also after, marriage as scientists, laboratory technicians, nurses, physicians, librarians, or artists. . . . My search for autistic children of unsophisticated parents has remained unsuccessful to date. . . . Most of the parents declare outright that they are not comfortable in the company of people; they prefer reading, writing, painting, making music, or just 'thinking' " (6, pp. 420–421). Rank, who also describes the use of intellectualization and obsessiveness in the parents of atypical children, reports that ". . . not too rarely they are highly intellectual, prominent people" (9, p. 131). Despert also emphasizes that mothers of these psychotic children seek gratification from intellectual sources rather than from contact with people. According to her, ". . . two significant characteristics of the mother and child which are frequently encountered in the group are, namely, the emotional detachment and the overintellectualization in the mother and the early absorption of the child in nonhuman content and nonfunctional activities, such as extensive acquisition of knowledge of musical records, license plate numbers, astronomical data, large and inappropriate vocabulary, word spelling, etc. On the one hand, the mother is remote and . . . on the other she puts excessive pressure on the growing baby to achieve feats of development which are considerably beyond his age level and interests" (4, p. 346).

The findings in the present research would, therefore, be predicted on the basis of these clinical reports. Discussion of similarities and differences between the present group of schizophrenics and those previously described as "autistic" or "atypical," as well as detailed description of the personalities of the mothers, are beyond the scope of the present report. It is only emphasized that for the kind of child diagnosed as schizophrenic at the Bradley Home, and for the kind of mothers of children who are admitted to this institution, the intelligence test findings fit well with previously reported observations.

Finally, the comparison between the IQs of the schizophrenic children and their nonschizophrenic emotionally disturbed peers who were also undergoing residential treatment is of interest. The schizophrenics, with an average IQ of only 84, score 8.7 IQ points below the other disturbed children. Since schizophrenic children are even more disturbed than the others, and since in most instances emotional disturbance detracts from intellectual functioning, this finding is in accord with the prediction of poorer performance by schizophrenics and, accordingly, is statistically significant.

In viewing these results of intelligence testing, it is noteworthy that the average IQ for this total sample is just about 92, which is considerably below the average for children in general. As mentioned above, some of the psychotic children are untestable, and as a group they perform very poorly. Even the ones who employ a mechanized, intellectualized approach to the world do not do very well on the standard tests because their specialized knowledge, obsessive concern for certain specific details, preoccupations with technical gadgets, astronomy, and so forth, does not earn them IQ points in the various areas of general knowledge that are tapped by conventional tests. Actually, the great store of anxiety with which they must cope and the kinds of defense mechanisms they employ make it impossible for them to benefit from contacts with people and events in their environment, and even the schizophrenic child who performs feats of memory in some particular area of concern or who seems to know a great deal about the physical universe or certain aspects of modern technology is unable to give an outstanding performance on intelligence tests. As for the over-all group of disturbed children, clinical psychologists who have tested them affirm that the majority are functioning considerably below their optimal level. Before being admitted to the Bradley Home, each child has to be evaluated as possessing the capacity for at least average intellectual functioning. That is, the psychologist must judge that even though the child tests subnormal (sometimes falling even into the "mental defective" category), basically, he is of normal or higher intelligence. The fact that as a group their intellectual functioning is greatly impaired is further demonstrated by the finding, mentioned above, that they perform so much worse than do their mothers. It seems unlikely that such discrepancies would be found between groups of normal children and their parents. In fact it seems highly plausible that, in many situations, groups of normal children would obtain higher IQs than their parents.

It is hoped that future studies of the personalities of these children and their mothers will contribute to further understanding of the dynamics of the psychopathology and nature of the intellectual deficit. It is evident from the present findings, however, that severe emotional disturbance has a pronounced detrimental effect on intellectual functioning in children. Undoubtedly, this state of affairs must set up a vicious circle, with disturbance leading to increasing intellectual failure and frustration, and the inability to achieve intellectual mastery leading to greater emotional maladjustment. In a residential treatment center, the therapeutic approach is from both points; with child care workers and recreational personnel helping the child to acquire increasing skills and abilities, special school teachers providing individualized and understanding teaching, and psychotherapists working with the child's emotional conflicts. The goal, of course, is to acquire emotional and intellectual light where now there is darkness. Objective knowledge of the intellectual and emotional fruits to be derived from these therapeutic efforts awaits much further investigation, but there is no doubt that these children are deserving of the scientific and clinical investment.

Summary

The purpose of this investigation was to study the intellectual functioning of 70 hospitalized emotionally disturbed children, of whom 12 were schizophrenic and 58 were suffering from other types of severe disturbance, and to compare the children's intelligence with a measure of their mothers' intelligence based on performance on the Wechsler-Bellevue scales. A significant positive correlation was found between the IQs of nonschizophrenic children and the IQs of their mothers, but no association was found between the IQs of the schizophrenic children and their mothers. Both groups of children received IQs significantly lower than did the groups of mothers. Although the mothers of the nonschizophrenic children received a mean IQ that was higher than the mean for the general population, the mean IQ of 122 for mothers of schizophrenic children was significantly higher. In comparing the two groups of children, it was found that the schizophrenics received significantly lower IQs, and that the mean IQ for this total sample of disturbed children fell at the bottom of the normal range of intelligence. Some theoretical, therapeutic, and practical implications of these findings were discussed.

Received June 27, 1957.

References

1. Anastasi, Anne, & Foley, J. P. *Differential psychology*. New York: Macmillan, 1949.
2. Burks, Barbara S. The relative influence of nature and nurture upon mental development: A comparative study of foster parent-foster child resemblance and true parent-true child resemblance. *Yearb. nat. Soc. Stud. Educ.*, 1928, **27**, Part I. Pp. 219–236.
3. Conrad, H. S., & Jones, H. E. A second study of familial resemblance in intelligence: environmental and genetic implications of parent-child and sibling correlations in the total sample. *Yearb. nat. Soc. Stud. Educ.*, 1940, **39**, Part II. Pp. 97–141.

4. Despert, J. Louise. Some considerations relating to the genesis of autistic behavior in children. *Amer. J. Orthopsychiat.*, 1951, **21**, 335–347.
5. Jones, H. E. The environment and mental development. In L. Carmichael (Ed.), *Manual of chilld psychology*. New York: Wiley, 1952. Pp. 631–696.
6. Kanner, L. Problems of nosology and psychodynamics of early infantile autism. *Amer. J. Orthopsychiat.*, 1949, **19**, 416–426.
7. Leahy, A. M. Nature-nurture and intelligence. *Genet. psychol. Monogr.*, 1935, **17**, 236–308.
8. Pearson, K. *Nature and nurture*. London: Dulau, 1910.
9. Rank, Beata. Adaptation of psychoanalytic technique for the treatment of young children with atypical development. *Amer. J. Orthopsychiat.*, 1949, **19**, 130–139.
10. Roff, M. A statistical study of the development of intelligence test performance. *J. Psychol.*, 1941, **11**, 371–386.
11. Thompson, G. G. *Child psychology*. Boston: Houghton Mifflin, 1952.

Journal of Consulting Psychology
Vol. 22, No. 3, 1958

Some Personality Correlates of Subjects Who Fail to Appear for Experimental Appointments[1]

Allan H. Frey and Wesley C. Becker
University of Illinois

In the course of an experiment using two extreme groups, introverts and extraverts, the characteristics of *S*s who failed to appear for an appointment with or without notifying the experimenter were investigated.

Five of the Guilford personality scales, Sociability, Rathymia, General activity, Social ascendance, and Emotional stability, were administered to a large section (250) of introductory psychology students. These tests were scored, and a group of introverts and a group of extraverts were selected by the experimenter on the basis of extreme Rathymia scores. Appointments were then made for each member of these groups by the use of postal cards.

Twelve *S*s notified the experimenter that they would be unable to keep the appointment (Group *N*); fourteen *S*s failed to appear without notifying the experimenter (Group *F*). The *N* group was characterized by low scores on scales *S, R, G,* and *A;* the *F* group by high scores. The chi-square test indicated that these differences are significant at the .01 level, except *R*, which is significant at the .001 level. There was no significant difference between groups on the *C* scale, nor was there a significant difference within or between groups with regard to sex. According to Eysenck, the *S, R, G,* and *A* scales are all highly loaded on a second-order factor which has been called extraversion–introversion.

The results of this study indicate that a sampling bias occurs when *S*s are obtained by the method noted above. This bias could have undesirable effects in many kinds of experiments. For example, in several recent studies, Eysenck has been attempting to relate extraversion–introversion to such variables as degree of satiation, conditionability, and reminiscence in pursuit rotor learning. Sampling bias of the kind indicated above would tend to attenuate any true relationships between these variables and extraversion–introversion. Of course, the generalizability of any findings is also affected. It is suspected that such bias would also enter most studies of conformity behavior, group dynamics, and anxiety. Anxiety studies enter the picture because of their correlation with extraversion–introversion. Franks reported an r of $-.48$ between Rathymia and the Taylor Anxiety scale. Farber *et al.* reported an r of $-.36$ between the same two tests. The present writers found a correlation of $-.32$ between Cattell's Anxiety scale and Eysenck's Extraversion scale. It therefore seems likely that selective sampling on extraversion would encompass to a lesser degree selective sampling on anxiety.

In practice, we have found it possible to reschedule the *N* subjects, but unless considerable pressure is exerted on the *F* subjects, they rarely appear when rescheduled. In experiments where extraversion–introversion is suspected as an important factor, it is suggested that pretesting on this variable be carried out while *S*s are in a "captive" state, e.g., in the classroom, and that special pressure be exerted to obtain cooperation from the extreme group members.

Brief Report.
Received January 2, 1958.

---

[1] An extended report of this study may be obtained without charge from Allan H. Frey, Department of Psychology, University of Illinois, Urbana, Ill., or for a fee from the American Documentation Institute. Order Document No. 5537, remitting $1.25 for microfilm or $1.25 for photocopies.

Journal of Consulting Psychology
Vol. 22, No. 3, 1958

Normative Data and Factor Analysis on the Parental Attitude Research Instrument[1]

Marvin Zuckerman and Beatrice Barrett Ribback
Indiana University Medical Center

Irwin Monashkin and James A. Norton, Jr.
Purdue University

The Parental Attitude Research Instrument (PARI), developed by Schaefer and Bell (3), consists of 23 five-item scales. Each scale measures an attitude which is theoretically relevant to personality development in children. The test items are modeled after the type of items developed by Mark (2) and Shoben (4) in their studies of parental attitudes. The items are general opinions with which the subject is asked to strongly or mildly agree or disagree on a four-point scale. An attempt was made to increase the social acceptability of items by putting them in the form of rationalizations, truisms, and cliches. The names of the 23 scales can be found in Table 4.

Schaefer and Bell's normative multiparae sample was 100 wives of military personnel who were given the test one to four days following delivery at a military hospital. There are two limitations to the representativeness of this sample: (*a*) the selectness of the occupational group and (*b*) the possibility that the expressed parental attitudes may have been influenced by the recency of parturition.

The senior author is engaged in a series of studies on parental attitudes in various groups of women including mothers of disturbed children. In collecting a normative sample we had four purposes: (*a*) to provide stanine scores for future work, (*b*) to investigate four possible influences on parental attitudes: age, education, number of children, and religion, (*c*) to obtain data for a factor analysis of the 23 scales, and (*d*) to provide a pool of control subjects for comparisons with other groups of mothers.

In order to facilitate interpretation in comparisons between different groups of mothers, it was deemed desirable to determine the factorial structure of the test. Schaefer and Bell analyzed the factorial structure of the test on their normative group of 100 mothers using the grouping method of factoring without rotation of axes. Their results revealed three factors in the test: Factor I included 16 of the 23 scales and seemed to be a controlling-authoritarian factor; Factor II included four scales and seemed to be a hostility-rejection factor; Factor III included three scales and seemed to be a democratic-equalitarian factor. The purpose of our factor analysis was to derive a single factorial structure from heterogeneous groups which would be useful in comparing factor scores across such groups; and to see if the factorial structure derived from such a heterogeneous sample would be similar to that obtained by Schaefer and Bell from a simple normal sample.

Procedure

Normative Data

The normative sample was composed of 222 mothers including mothers of nursery school children, mothers of young adults attending a night extension college, and mothers

---

[1] The authors would like to thank Donald F. Moore, Superintendent of Carter Memorial Hospital, for his kind permission to use the facilities of Carter Hospital to collect data from patients and their mothers. A note of thanks is extended to James Besheers for his assistance in the mechanics of the factor analysis.

Bureau of Ednl. Research
DAVID HA
NG COLLEGE
No

contacted at meetings of church social groups. Relevant data on this sample appears in Tables 1–3. Using church social groups gave us the advantage of being able to secure "captive" groups, thus avoiding the selective effects entailed in using volunteers. However, use of these groups may introduce another selective factor since they comprise slightly over one-half of the normative sample. In addition, it should be noted that the sample is drawn exclusively from an urban, midwestern population. It contains a greater representation of women with some college education and a greater proportion of older mothers than is found in the general population. The religious and racial proportions of the sample are not too deviant from those of the general population.

Factor Analysis

The following subjects were added to the normal group of 222 to comprise the sample for the factor analysis: (*a*) 60 mothers who were psychiatric patients in an acute treatment hospital, (*b*) 69 mothers of young adult psychiatric inpatients, and (*c*) 62 mothers of children referred for examination to two child guidance clinics and a children's inpatient psychiatric service. The total *N* for the factor analysis was therefore 413. The mean education of this sample was 11.9 years, *SD* 2.9. The mean age was 38.6, *SD* 10.1.

Table 1

Education Distribution of Normative Sample
($M = 12.73 \quad SD = 2.79$)

Grades	*N*	%
0–6	4	2
7	5	2
8	13	6
9	6	2
10	6	2
11	9	4
12	79	36
13	17	8
14	22	10
15	11	5
16	44	20
17	2	1
18	2	1
19	1	1
20	1	1
	222	99

Table 2

Age Distribution of Normative Sample
($M = 37.94 \quad SD = 8.71$)

Age	*N*	%
0–19	1	0
20–24	5	2
25–29	25	11
30–34	59	27
35–39	47	21
40–44	40	18
45–49	24	11
50–54	9	4
55–59	7	3
60–64	3	1
65–70	2	1
	222	98

Scores of all 413 subjects were punched on cards and intercorrelated on the 650 IBM machine at Indiana University. The largest absolute value of the correlations in each column of the matrix was used as the communality estimate for the scale. The matrix was factored, using the principal axes solution, on Purdue University's Datatron computer. The first three principal components accounted for a total of 11.96 out of an estimated communality variance of 13.57. It was decided, tentatively, to stop with three factors, since the principal components after the third (which were also computed) added only

Table 3

Number of Children, Distribution in
Normative Sample
($M = 2.53 \quad SD = 1.46$)

No.	*N*	%
0	6	3
1	46	21
2	76	34
3	48	22
4	28	12
5	5	2
6	11	5
8	1	0
9	1	0
	222	99

Table 4

Means, *SD*s, and Stanine Scores for the PARI

($N = 222$)

Scale		9	8	7	6	5	4	3	2	1	Mean	*SD*
1	Encouraging Verbalization	5–14	15	16	17	18	19	20	—	—	17.93	1.95
2	Fostering Dependency	20–17	16–15	14	13–12	11	10–9	8–7	6	5	10.82	3.08
3	Seclusion of the Mother	20–18	17–16	15–14	13	12–11	10–9	8	7–6	5	11.39	3.38
4	Breaking the Will	20–17	16–15	14	13–12	11–10	9	8–7	6–5	—	10.54	3.29
5	Martyrdom	20–16	15–14	13	12–11	10–9	8–7	6	5	—	9.44	3.54
6	Fear of Harming the Baby	20–19	18	17–16	15–14	13–12	11	10–9	8–7	6–5	12.81	3.40
7	Marital Conflict	—	20	19–18	17–16	15	14–13	12–11	10	9–5	15.04	3.33
8	Strictness	20	19–18	17	16–15	14–13	12	11–10	9	8–5	13.71	3.22
9	Irritability	20	19–18	17	16–15	14	13	12–11	10	9–5	14.22	2.81
10	Excluding Outside Influences	20–18	17–16	15–14	13–12	11	10–9	8–7	6–5	—	10.90	3.54
11	Deification	20	19–18	17–16	15–14	13–12	11	10–9	8–7	6–5	12.80	3.69
12	Suppression of Aggression	20–15	14	13	12–11	10	9–8	7	6	5	10.03	2.75
13	Rejection of Homemaking Role	20–18	17–16	15	14–13	12–11	10	9–8	7–6	5	11.62	3.23
14	Equalitarianism	5–13	14	15	16	17	18	19	20	—	17.40	2.03
15	Approval of Activity	20–19	18–17	16	15–14	13	12–11	10	9–8	7–5	12.84	3.02
16	Avoidance of Communication	20–17	16–15	14	13–12	11	10–9	8–7	6	5	10.81	3.20
17	Inconsiderateness of Husband	20–19	18–17	16–15	14–13	12	11–10	9–8	7–6	5	11.96	3.61
18	Suppression of Sex	20–14	13	12–11	10	9–8	7	6–5	—	—	8.71	2.97
19	Ascendance of the Mother	20–19	18–17	16–15	14–13	12–11	10	9–8	7–6	5	11.83	3.55
20	Intrusiveness	20–17	16–15	14	13–12	11–10	9–8	7–6	5	—	10.39	3.51
21	Comradeship and Sharing	5–15	16	17	18	19	—	20	—	—	18.67	1.70
22	Acceleration of Development	20–18	17	16–15	14–13	12–11	10–9	8	7–6	5	11.67	3.56
23	Dependency of the Mother	20–18	17	16–15	14–13	12	11–10	9–8	7	6–5	12.00	3.24

small amounts to the total of explained variance. This decision was confirmed by examination of the residuals after the three factors: of 276 residuals, only seven were over .10, the largest being .16. It was felt that this was well within the sampling variation of the original correlations (for 400 *df*, the 5% level for significance of difference is .098). Orthogonal rotations were carried out on the factor matrix using the single plane method (actually only one rotation in the plane of the second and third components seemed indicated).

Results and Discussion

Normative Data

The scales and their means, standard deviations and stanine scores derived from the normal group, are given in Table 4. The correlations of the scales with age, education, and number of children are given in Table 5. The education of the mother seems to be the most significant variable of the three, relating to parental attitudes. Using the one per cent level for significance, education correlates significantly with 18 of the 23 scales, age with six of the scales, and number of children with none of the scales. The direction of the correlations indicate that the less educated and older mothers tend to have more severe attitudes (e.g., more authoritarian, suppressive, hostile) on the scales which are related.

Samples of 38, equivalent in age and education, were taken from each of two religious groups, Catholic and Protestant. Comparisons of the means on the individual scales were made using simple analysis of variance. A significant difference was found only on Scale 7, Marital Conflict, on which the Catholics scored significantly higher. However, one difference out of 20 tests might be expected by chance; therefore, we must not conclude that religion (Catholic and Protestant) is a significant influence on parental attitudes as measured by the PARI. It might be noted, however, that there exists a rather homogeneous culture in this area. Differences between religious groups may be greater in sections of the country where there is less homogeneity of culture. The size of the samples also limits any broad conclusions drawn from these data.

Factor Analysis

The correlation matrix of the 23 PARI scales, education and age, based on the total *N* of 413, can be seen in Table 6. The figures in the diagonals are the communality estimates. The unrotated factor matrix can be seen in Table 7 and the rotated factor matrix in Table 8.

The three factors extracted in this study are very similar to those found by Schaefer and Bell, with some minor exceptions which will be noted in the discussion of Factor B. Thus, the factor structure over a clinically heterogeneous group of mothers appears to be substantially the same as that obtained on normal mothers only.

Factor A has loadings of .60 and above on 16 of the 23 scales. The significant loadings are on the same scales as found in Schaefer and Bell's Factor I. The loadings are generally higher in this study, probably reflecting the greater heterogeneity of our sample. The scales which contribute most heavily to this factor are those which measure authoritarian, suppressive, punitive, and restricting types of attitudes. However, this factor is so broad that it may be measuring a general response set to the test as well as specific response to the content of the scales. Education has a negative loading on this factor indicating that the more highly educated mothers are more likely to score low on this factor.

Factor B is similar to Factor II in Schaefer and Bell's analysis except that Scale 8, Strictness, no longer is part of the factor. Factor B is defined mainly by the two scales, 9 and 13, Irritability and Rejection of the Homemaking Role. Scale 7, Marital Conflict, shows a smaller loading on this factor. The character of this factor can be described as hostility toward children and husband and rejection of the maternal role. Unlike Factor A, this fac-

Table 5

Correlations Between PARI Scales and Age, Education, and Number of Children[a]

Scale		Education	Age	No. of children
1	Encouraging Verbalization	.15	.05	−.20
2	Fostering Dependency	−.34*	.11	.07
3	Seclusion of the Mother	−.47*	.14	.19
4	Breaking the Will	−.41*	.13	.14
5	Martyrdom	−.41*	.09	.04
6	Fear of Harming the Baby	−.31*	.08	−.03
7	Marital Conflict	−.21*	.04	.18
8	Strictness	−.25*	.17*	.11
9	Irritability	−.08	−.14	.23
10	Excluding Outside Influences	−.46*	.16	.19
11	Deification	−.37*	.01	.18
12	Suppression of Aggression	−.27*	.02	.07
13	Rejection of Homemaking Role	−.06	−.04	.03
14	Equalitarianism	.12	.11	−.05
15	Approval of Activity	−.26*	.17*	.10
16	Avoidance of Communication	−.39*	.29*	.02
17	Inconsiderateness of Husband	−.29*	.18*	.01
18	Suppression of Sex	−.43*	.22*	.00
19	Ascendance of the Mother	−.32*	.12	.08
20	Intrusiveness	−.31*	.00	.00
21	Comradeship and Sharing	−.01	.02	.10
22	Acceleration of Development	−.42*	.23*	.10
23	Dependency of the Mother	−.29*	.01	.00

* Significant at or below the .01 level.

[a] Number of children was not available for part of the normative sample at the time the correlations were run. The correlations given for age and education are based on an *N* of 222. Correlations for No. of Children are based on an *N* of 125.

Table 6

Correlation Matrix; 23 PARI Scales, Education, and Age

Scale	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	Ed.	Age
1	(36)	−07	−13	−13	−19	−05	04	−10	06	−21	−18	−11	00	36	−07	−15	−10	−23	−12	−20	28	−13	−08	13	01
2		(59)	55	53	57	46	25	33	05	59	53	52	15	02	44	43	40	58	49	55	13	55	50	−45	25
3			(62)	54	59	50	31	39	06	62	50	48	12	11	52	47	51	62	56	55	08	60	47	−53	20
4				(65)	65	45	36	48	26	62	60	48	27	01	49	54	46	59	54	55	10	56	48	−47	18
5					(65)	50	32	40	19	60	55	53	27	−03	48	49	50	64	53	58	00	53	51	−44	13
6						(53)	31	26	32	50	42	38	35	16	45	46	43	53	51	46	10	46	51	−35	10
7							(38)	23	33	24	23	10	26	16	30	29	38	20	31	17	12	28	26	−22	01
8								(50)	15	37	48	29	16	08	50	38	27	33	43	40	19	38	34	−21	20
9									(51)	09	12	03	51	15	12	24	21	−01	18	01	06	05	23	02	−08
10										(70)	70	57	16	02	50	58	48	65	54	62	10	56	53	−51	20
11											(70)	48	09	−03	48	49	35	58	53	59	18	52	48	−41	14
12												(57)	22	−01	43	44	37	57	40	53	04	47	44	−39	19
13													(51)	17	20	27	34	12	26	12	05	14	31	−03	02
14														(37)	15	05	11	−06	09	−04	37	09	08	06	02
15															(59)	56	46	47	59	55	14	54	44	−31	22
16																(58)	43	54	50	48	05	49	40	−37	34
17																	(51)	44	49	38	04	49	47	−32	21
18																		(65)	55	60	−01	61	49	−50	29
19																			(59)	57	14	46	42	−39	21
20																				(62)	08	55	46	−37	12
21																					(37)	13	11	−05	06
22																						(61)	49	−48	32
23																							(53)	−33	10
Ed.																								(53)	−23
Age																									(34)

Table 7

Unrotated Factor Matrix

Scale No.	Name	Factor I	Factor II	Factor III	h²
1	Encouraging Verbalization	−.18	.37	−.40	.33
2	Fostering Dependency	.71	−.10	−.11	.53
3	Seclusion of the Mother	.76	−.08	−.10	.59
4	Breaking the Will	.77	.04	.08	.61
5	Martyrdom	.77	−.04	.18	.63
6	Fear of Harming the Baby	.66	.22	.09	.48
7	Marital Conflict	.39	.36	.07	.29
8	Strictness	.54	.09	−.13	.31
9	Irritability	.21	.60	.34	.52
10	Excluding Outside Influences	.80	−.15	−.01	.67
11	Deification	.74	−.12	−.06	.56
12	Suppression of Aggression	.65	−.15	.00	.45
13	Rejection of the Homemaking Role	.31	.53	.33	.48
14	Equalitarianism	.07	.49	−.40	.40
15	Approval of Activity	.70	.10	−.13	.51
16	Avoidance of Communication	.70	.06	.05	.49
17	Inconsiderateness of the Husband	.63	.17	.10	.43
18	Suppression of Sex	.78	−.26	.03	.67
19	Ascendance of the Mother	.72	.08	−.02	.53
20	Intrusiveness	.73	−.19	−.02	.58
21	Comradeship and Sharing	.14	.31	−.49	.35
22	Acceleration of Development	.74	−.08	−.14	.57
23	Dependency of the Mother	.66	.13	.07	.45
24	Education	−.58	.22	.07	.38
25	Age	.28	−.11	−.22	.14
Sum Aij^2		9.37	1.61	0.98	11.97 11.96

tor seems to be unrelated to the education of the mother.

Factor C is identical to Factor III in the Schaefer and Bell analysis and consists of three scales: 1, Encouraging Verbalization; 14, Equalitarianism; and 21, Comradeship and Sharing. Factor C can be described as measuring democratic attitudes toward child rearing. Theoretically, one would not expect to find this as an independent factor since these scales should load on the negative side of Factor A, Authoritarianism. Apparently, some mothers score high on both of these factors which are somewhat contradictory in content. Perhaps the acquiescent response set, discovered in the F scale by Bass (1), may be responsible for this result. Many mothers may agree with contradictory statements because of the tendency to agree with any sort of positive sounding generality. Acquiescent response set on the PARI is currently being investigated by the senior author.[2]

Summary

Normative data on the Parental Attitude Research Instrument were gathered from 222 mothers. Stanine scores for the 23 scales were derived from the data. Correlations were done between the scales and the education, age, and

---

[2] Subsequent to this factor analysis of the total heterogeneous sample, separate factor analyses were carried out on the group of 131 mothers of disturbed children and adults and on the group of 222 mothers of normals. Although the correlations between scales were higher in the group of mothers of disturbed children and adults than in the mothers of normals, Factors A, B, and C emerged from each analysis and were defined by essentially the same scales as reported in the analysis of the total sample. These results indicate that the obtained factors are stable over two different samples and may be used for comparisons between similar criterion groups.

Table 8

Rotated Factor Matrix

Scale No.	Name	A	B	C	h²
1	Encouraging Verbalization	−.18	.07	.55	.34
2	Fostering Dependency	.71	−.15	.03	.53
3	Seclusion of the Mother	.76	−.12	.03	.59
4	Breaking the Will	.77	.08	−.04	.61
5	Martyrdom	.77	.07	−.17	.63
6	Fear of Harming the Baby	.66	.23	.05	.48
7	Marital Conflict	.39	.34	.15	.29
8	Strictness	.54	.00	.15	.31
9	Irritability	.21	.69	.07	.52
10	Excluding Outside Influences	.80	−.12	−.08	.67
11	Deification	.74	−.14	−.02	.56
12	Suppression of Aggression	.65	−.12	−.09	.45
13	Rejection of the Homemaking Role	.31	.62	.03	.48
14	Equalitarianism	.07	.17	.60	.40
15	Approval of Activity	.70	.00	.16	.51
16	Avoidance of Communication	.70	.08	−.01	.49
17	Inconsiderateness of the Husband	.63	.20	.02	.43
18	Suppression of Sex	.78	−.20	−.17	.67
19	Ascendance of the Mother	.72	.06	.07	.53
20	Intrusiveness	.73	−.17	−.09	.58
21	Comradeship and Sharing	.14	−.03	.58	.35
22	Acceleration of Development	.74	−.15	.07	.57
23	Dependency of the Mother	.66	.14	.01	.45
24	Education	−.58	.21	.07	.38
25	Age	.28	−.21	.11	.14
					11.97
Sum Aij²		9.37	1.40	1.20	11.97

number of children of the mother. The education of the mother was most significantly related to scores on the scales. Age showed a smaller number of relationships, and number of children no significant relationships with the scales.

The scales were factor analyzed using a heterogeneous sample composed of the 222 mothers in the normative sample, plus 131 mothers of psychiatric patients and disturbed children and 60 mothers who were psychiatric patients. Three factors were extracted from the test:

1. Factor A, Authoritarian-Control
2. Factor B, Hostility-Rejection
3. Factor C, Democratic Attitudes

The influence of response set on the test was considered.

Received July 1, 1957.

References

1. Bass, B. M. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.,* 1955, **51**, 616–623.
2. Mark, J. C. The attitudes of the mothers of male schizophrenics toward child behavior. *J. abnorm. soc. Psychol.,* 1953, **48**, 185–189.
3. Schaefer, E. S., & Bell, R. Q. Development of the Parental Attitude Research Instrument. *Child Developm.,* in press.
4. Shoben, E. J. The assessment of parental attitudes in relation to child adjustment. *Genet. Psychol. Monogr.,* 1949, **39**, 101–148.

Journal of Consulting Psychology
Vol. 22, No. 3, 1958

A Self-Scored Projective Measure of Achievement and Affiliation Motivation[1]

Richard deCharms
Washington University

The measure of motivation recently developed by McClelland *et al.* (1) for experimental purposes has the disadvantage of requiring a careful content analysis of projective materials. The procedure to be described here was developed with the aim of obtaining quickly a motive score that is as close to the score derived from content analysis as possible.

Each *S* wrote six stories in response to six picture stimuli under the standardized conditions described by McClelland *et al.* (1). *S* was then given a form containing several sample story plots for each stimulus picture and asked to rank order their similarity to his own story. A score was derived from this form by totalling the rankings of motive relevant plots on the latter form. These "self-scored" motive scores were compared with the scores derived from content analysis of the original stories written by *S*.

Achievement motivation. Scores for three groups of *S*s (total $N = 183$), were obtained from the self-scored measure by totalling the rankings of achievement related plots. Analysis of variance of the means of the content analysis measure classified as high, medium, and low by the self-scored measure demonstrated that the means differed significantly in all three groups ($p < .001$). For a fourth group ($N = 47$), the self-scored measure of achievement motivation discriminated significantly between volunteers and nonvolunteers for a psychological experiment. The content analysis score has previously been shown to be related to volunteering.

Affiliation motivation. Scores for three groups (total $N = 128$) were obtained from the self-scored measure by totalling the rankings of the affiliation related plots. Analysis of variance of the means of the content analysis measure classified as high, medium, and low by the self-scored affiliation measure demonstrated that the means differed significantly in all three groups ($p < .025$).

The self-scoring procedure should not be as sensitive and reliable as actual content analysis. It will, however, yield an immediate score which could be used to identify *S*s with high or low motive scores, and will be especially valuable where the test could be administered to a large number of *S*s from which extremes could be drawn. Since much of the work of scoring has been eliminated, the new measure will allow the researcher to use more *S*s or to replicate results.

Brief Report.
Received December 14, 1957.

Reference

1. McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

---

[1] An extended report of this study may be obtained without charge from Richard deCharms, Graduate Institute of Education, Washington University, St. Louis 5, Missouri, or for a fee from the American Documentation Institute. Order Document No. 5536, remitting $1.75 for microfilm or $2.50 for photocopies.

Journal of Consulting Psychology
Vol. 22, No. 3, 1958

Asynchronism and Apparent Movement Thresholds in Brain-Injured Children[1]

Henry J. Mark
Department of Otolaryngology, Johns Hopkins University

and Benjamin Pasamanick
Department of Psychiatry, Ohio State University

This paper reports some differences in foveal asynchronism and apparent movement thresholds, and peripheral two-point and apparent movement thresholds, found between brain-injured (pyramidal tract damaged) and control children. The results were obtained in the course of a larger study aimed at discovering neuropsychological measurements which might be used to detect otherwise unnoticed alterations in cerebral functioning. Previous studies by Klein and Krech (4) and Eysenck (3) demonstrated altered figural after-effects in brain injury, and studies with the same subjects (*S*s) as those reported here with critical flicker fusion (CFF) (8) and light perception (7) threshold measurements suggest that visual threshold measurements may be sensitive to the effect of nonoccipital lobe lesions.

Asynchronism threshold measurements were selected for study here to determine whether lowered visual efficiency and increased intraindividual variability in the brain-injured could again be demonstrated as in the previous studies with these groups (7, 8). Apparent movement threshold measurements were selected for study because an explanation of the phenomenon of apparent movement usually invokes one form or another of a stimulus after-effect or stimulus surround-effect theory. By presenting binocular stimuli in homonymous half fields, Birch and Ashdourian (2) demonstrated that central mechanisms by themselves are capable of producing an apparent movement (phi) phenomenon. Since central mechanisms are involved, such mechanisms may then give rise to differences in apparent movement thresholds between brain-injured and control groups. In addition, peripheral mechanisms such as retinal facilitation (9, p. 140), which are known to contribute to the perception of the apparent movement phenomenon, may well be altered in children whose brain injury has been present since birth. This may be hypothesized on the basis of epidemiologic studies (5, 6, 10, 11, 12) which point to anoxia (rather than traumatic injuries) as the major cause of brain damage. Since the neurons of the retina are also susceptible to damage from anoxia, altered retinal stimulus after-effects or surround-effects may combine with central mechanisms to produce apparent movement thresholds which distinguish brain-injured from control *S*s.

Hypotheses

Higher asynchronism thresholds, corresponding to loss of visual efficiency in the detection of a pause (p) between the onset time of two lights, were predicted in the brain-injured. With p as the variable, apparent movement thresholds and peripheral two-point thresholds cannot be readily interpreted in terms of visual efficiency, and the direction of the difference between brain-injured and control *S*s in these thresholds was not specified in our hypothesis. Greater intra-individual variability, reported as the *V* score, was predicted in the brain-injured. It was also hoped that differential reactions to peripheral as compared to foveal measurements would again provide good discriminators between brain-injured and control *S*s, as it did in two CFF experiments (7, 8).

---

[1] This research was supported by the Alfred P. Sloan Foundation, Inc. grant to the Division of Laryngology and Otology, Department of Surgery, Johns Hopkins University, as well as by a grant from the Foundation for Mentally Retarded and Handicapped Children of Baltimore. The authors wish to thank Paul Meier for invaluable assistance and advice, and both Bernard Weiss and Victor Laties for their critical reading of the manuscript.

The Experiment

Subjects

The brain-injured group was composed of ten children (six males, four females) with diagnosed pyramidal tract involvement and with no primary diagnosed opthalmologic pathology. The control group was composed of ten children (five males, five females) with conditions which did not justify an assumption of brain damage or injury such as spina bifida, hemophilia, or congenital amputation. All *S*s attended a Baltimore school for handicapped children. The groups were roughly comparable in terms of age, IQ, and sex, and the effects of these covariates on all significant discriminators were examined by means of regression diagrams. None of these was significantly related to performance. For example, there was no apparent regression of thresholds or *V* scores on IQ (which ranged between 64 and 121), and it was therefore assumed that IQ did not enter as an important experimental factor in the difference between brain-injured and control groups. The mean IQ of the brain-injured group was 89, with an *SD* of 16; the mean for the control group was 96, with an *SD* of 16. The age range was between 10 and 15 years, with a median of 12 for both groups.

Apparatus

Two Sylvania R1131 C glow-modulator tubes were mounted centrally on a perimeter so that two rectangular focused lines of light, each ¾-mm wide and 6-mm long, and separated by a distance of 4 mm, could be projected by means of a simple achromatic lens onto a piece of white plastic mounted on black velvet cloth. The glow modulators were activated by an electronic pulse-type stimulator and timer (13) which permitted continuous variation from 4 ms to 260 ms, of the pause (p) between the onset of the two light stimuli. The lights were viewed by the *S*s from a distance of 33 cm. The two lights were matched for wave length and had a brightness of the order of one-tenth millilambert (0.1 ml) as determined with a MacBeth Illuminometer. The light stimuli were presented once per second with the duration of both stimulus lights held constant at 135 ms. For peripheral threshold determinations, a 5-mm diameter circle of luminous paint served as a fixation point. The experiments were carried out in total darkness.

Procedure

The first binocular foveal measurement was always an ascending asynchronism threshold (A_{Asc}) measurement, where the pause (p) between the onset of the two light stimuli was continuously increased from 4 ms until *S* reported that the two lights were no longer presented simultaneously. As p was further increased, all *S*s noted movement effects, and the ascending apparent movement threshold (M_{Asc}) was recorded when *S* reported that the observed phenomenon appeared equivalent to the real movement of one line moving from left to right. Similarly, descending apparent movement (M_{Desc}) and asynchronism threshold (A_{Desc}) measurements, which may be considered simultaneity thresholds, were obtained. In this way, six ascending and six descending threshold measurements were obtained foveally. These were followed by three left-eye peripheral threshold measurements. With p at 4 ms, each *S* was able to distinguish two separate lines of light at 15 degrees nasal while fixating the 5-mm luminous circle. As p was increased, the two lines became blurred into a single vaguely defined area, and the ascending two-point threshold (TP_{Asc}) was recorded when *S* reported he was no longer able to distinguish two separate lines. As p was further increased, a left-to-right apparent movement of, and within, the vaguely defined area could be observed, and the peripheral apparent movement threshold (M_{P}) was recorded when *S* first reported such a movement. Then p was decreased until *S* again reported two separate lines. This was taken as the descending two-point threshold (TP_{Desc}).

The instructions, aimed at minimizing differences between ascending and descending thresholds, follow:

> For A_{Asc}: Now you see two lines of light going on and off at exactly the same time, exactly together. Let me know as soon as you notice that they are not going on exactly together any more, let me know as soon as you notice that they are out of order even the slightest bit.
>
> For M_{Asc}: Now you see two lines out of order, sort of chasing each other. Let me know as soon as you cannot see two lines any more. Let me know as soon as it looks just like one line that's moving across from left to right.
>
> For M_{Desc}: Now you see only one line that is moving from left to right. Let me know as soon as you can tell there are two lines again, as soon as you can tell it is not just one line that is moving from left to right.
>
> For A_{Desc}: Now you see two lines blinking out of order. Let me know as soon as you notice that they are going on and off exactly together, that they are not out of order even the slightest bit.
>
> For TP_{Asc}: Keep looking at the blue light (the luminous fixation point). Let me know as soon as you cannot make out two separate lines any more, let me know as soon as it looks just like one blotch of light going on and off. Don't take your eyes off the blue light.
>
> For M_P: Keep on looking at the blue light. Let me know as soon as the blotch looks like it is moving a little from left to right, as soon as it doesn't just seem to go on and off, as soon as it seems to move across a little bit. Don't take your eyes off the blue light.
>
> For TP_{Desc}: Keep on looking at the blue light. Let me know as soon as it no longer looks like just a blotch, but as soon as you can make out two separate lines, or two distinct lights again. Don't take your eyes off the blue light.

The estimated threshold for each *S* in this experiment is always the mean of six threshold measurements. The *V* score is always taken as the logarithm of the variance of the distribution [2] of measurements around the mean, i.e., around the threshold. Nonparametric rank methods (14) were indicated in view of the sometimes erratic estimates of spread (*V* scores) of the psychometric functions. The .05 level of significance was adopted as the criterion for rejecting the null hypothesis of no difference between groups.

Results and Discussion

Higher average thresholds were found in the brain-injured than in the control group (Table 1). For ascending asynchronism thresholds, it required an average pause of 44.5 ms for the brain-injured to notice that the two lights were not presented simultaneously, while an average pause of only 34.0 ms between the onset of the two lights was noticed by the control group. The difference of 10.5 ms between the groups is significant at the .05 level and supports the hypothesis that visual efficiency is impaired in the brain-injured. The average pause between the two lights for the ascending apparent movement threshold was 110.8 ms for the brain-injured group and 83.9 ms for the controls; the difference of 26.9 ms between the two groups is significant at the .01 level.

Table 1

Average Foveal Asynchronism Thresholds (A_{Asc} and A_{Desc}), Foveal Apparent Movement Thresholds (M_{Asc} and M_{Desc}), Peripheral Two-Point Thresholds (TP_{Asc} and TP_{Desc}), Peripheral Movement Thresholds (M_P), and Foveal-Minus-Peripheral Movement Scores ($M_{\mathrm{Asc}} - M_P$) in Milliseconds for the Brain-Injured and Control Groups of Ten *S*s Each

	Brain-injured		Control		
Thresholds	$\bar{X}$	*SD*	$\bar{X}$	*SD*	Diff.
Foveal					
A_{Asc}	44.5	11.2	34.0	8.1	10.5*
A_{Desc}	33.3	9.6	30.0	4.9	3.3
M_{Asc}	110.8	20.0	83.9	20.7	26.9**
M_{Desc}	81.2	10.9	72.1	15.6	9.1
Peripheral					
TP_{Asc}	53.7	17.5	52.8	19.5	.9
TP_{Desc}	53.6	10.6	41.6	19.3	8.0
M_P	92.7	22.0	97.3	32.0	4.6
Foveal-minus-peripheral					
$M_{\mathrm{Asc}} - M_P$	18.1	17.8	−13.4	28.0	31.5**

* Significant at the .05 level of confidence.
** Significant at the .01 level of confidence.

---

[2] A useful property of such a distribution of log variance scores is that their variance will depend *only* on the degrees of freedom in the estimate of the *V* score (1). If the psychometric function is a cumulative normal curve, then the variance of $\log_{10} s^2$ with 5 *df* is .0938 (if the curve is not a grossly non-normal cumulative curve, this is still approximately true). This makes it possible to compare the variance obtained for any of the average *V* scores in Table 2 with the known constant variance of .0938 by the *F* ratio and establish the significance level.

Table 2

Average *V* Scores (Log Variance Scores) for Foveal Asynchronism (A_{Asc} and A_{Desc}), Foveal Apparent Movement (M_{Asc} and M_{Desc}), Peripheral Two-Point (TP_{Asc} and TP_{Desc}), and Peripheral Apparent Movement Measurements for the Brain-Injured and Control Groups of Ten *S*s Each

Thresholds	Brain-injured $\overline{X}$	Brain-injured F^a	Control $\overline{X}$	Control F^a
Foveal				
A_{Asc}	2.05	2.62	1.69	2.10
A_{Desc}	2.19**	.42	1.74	1.44
M_{Asc}	2.16	5.03	2.11	1.96
M_{Desc}	2.02	1.55	2.06	2.11
Peripheral				
TP_{Asc}	2.44**	.78	1.98	1.67
TP_{Desc}	2.49	1.64	2.21	.69
M_P	2.60	1.26	2.23	1.78

** Significant at the .01 level of confidence.

[a] If in fact all *S*s within a group had the same intra-individual variability for a particular threshold, then $F_{(9,\infty)} \geq 1.88$ would occur only 1 time in 20.

When each *S*'s peripheral movement threshold (M_P) is compared with (i.e., subtracted from) his own corresponding foveal movement threshold (M_{Asc}), the brain-injured decreased their thresholds by an average of 18.1 ms, while the controls increased their thresholds by 13.4 ms. The difference between the two groups in this respect is significant at the .01 level and suggests that just as in previous CFF experiments, changes in thresholds due to altered conditions may yield good discriminators between the two groups.

Six out of seven average *V* scores of the brain-injured group were larger than those of the control group (Table 2). The *V* scores for foveal descending asynchronism thresholds and peripheral ascending two-point thresholds discriminated between the two groups at the .01 level of significance. Such findings throughout these studies support one of the major hypotheses that the brain-injured show increased intra-individual variability.

The visual thresholds in this experiment on the whole did not discriminate as sharply between brain-injured and control groups as did the light perception thresholds of the previous investigation (7). In view of the fact that all four *V* scores associated with the light perception thresholds discriminated between the two groups at the .01 level, the findings here serve to strengthen the major hypotheses but hold no additional promise as potential diagnostic discriminators. They may be of some theoretical interest however. Differences in apparent movement thresholds in the previously reported experiment, where the exposure time of the left light was the only experimental variable, are readily attributable to difference in exposure time required for light perception thresholds between the two groups. In this study where the pause between the two lights was the only experimental variable altered, central mechanisms may well have contributed to differences between the groups. Klein and Krech (4) and Eysenck (3) suggest that distinct nosological groups have characteristic differences in central after-effects. Experiments designed to test specific figural after-effect hypotheses related to magnitude, speed of arousal, and persistence of after-effects are reported elsewhere (7).

Summary

Differences in asynchronism, apparent movement, and peripheral two-point thresholds found between brain-injured (pyramidal tract damaged) and control children support the two major hypotheses of impairment in visual efficiency, and increased intra-individual variability in the brain-injured. The possible significance of the apparent movement threshold findings is discussed.

Received July 15, 1957.

References

1. Bartlett, M. S., & Kendall, D. G. The statistical analysis of variance-heterogeneity and the logarithmic transformation. *J. Roy. statist. Soc.* (*Suppl.*), 1946, **8**, 128–138.
2. Birch, H. G. Personal communication. 1955.
3. Eysenck, H. J. Cortical inhibition, figural after-effect, and theory of personality. *J. abnorm. soc. Psychol.*, 1955, **51**, 94–106.
4. Klein, G. S., & Krech, D. Cortical conductivity in the brain-injured. *J. Pers.*, 1952, **21**, 118–148.
5. Lilienfeld, A. M., & Pasamanick, B. Association of maternal and fetal factors with the development of epilepsy: I. Abnormalities in the prenatal and paranatal periods. *J. Amer. med. Ass.*, 1954, **155**, 719–724.

6. Lilienfeld, A. M., & Pasamanick, B. The association of prenatal and paranatal factors with the development of cerebral palsy and epilepsy. *Amer. J. Obstet. Gynecol.,* 1955, **70**, 93–101.
7. Mark, H. J. Studies of perception in brain-injured children. Unpublished doctoral dissertation, Johns Hopkins University, School of Hygiene and Public Health, 1955.
8. Mark, H. J., Meier, P., & Pasamanick, B. Variability of critical flicker fusion thresholds in brain-injured children. *A.M.A. Arch. Neurol. Psychiat.,* in press.
9. Morgan, C. T., & Stellar, E. *Physiological psychology.* New York: McGraw-Hill, 1950.
10. Pasamanick, B., & Lilienfeld, A. M. Maternal and fetal factors in the development of epilepsy: II. Relationship to some clinical features of epilepsy. *Neurology,* 1955, **5**, 77–83.
11. Pasamanick, B., & Lilienfeld, A. M. The association of maternal and fetal factors with the development of mental deficiency: I. Abnormalities in the prenatal and paranatal periods. *J. Amer. med. Ass.,* 1955, **159**, 155–160.
12. Rogers, M. E., Pasamanick, B., & Lilienfeld, A. M. Prenatal and paranatal factors in the development of childhood behavior disorders. *Acta psychiat., Kbh.,* 1955, Suppl., **102**, 157.
13. Roush, R. G., & Urbanski, E. T. Universal medical timer and pulse stimulator. *Electronics,* 1953, **26** (II), 154–157.
14. Wilcoxon, F. *Some rapid approximate statistical procedures.* New York: American Cyanamid, 1949.

Journal of Consulting Psychology
Vol. 22, No. 3, 1958

Opinions of Psychologists and Psychiatrists About Mental Health Problems[1]

Jum Nunnally
University of Illinois

The research reported here grew out of a larger program concerning the communication of information about mental health to the general public.[2] The first step in this larger program was to learn the state of public understanding about mental health problems. In order to do this, it was necessary to learn what the public thinks about mental health problems and to compare this with the opinions of experts in the mental health field. The studies of public opinion are reported elsewhere (1). This paper presents results obtained from a study of experts.

Questionnaire Construction

The term "opinions" refers in the specific sense to segments of verifiable knowledge. For example, the statement, "There are more men than women in mental hospitals," is classified as an opinion, because it could be verified one way or the other. On the other hand, a statement such as, "I am afraid to be around anyone who has had a nervous breakdown," indicates a personal reaction, or attitude. The studies of attitude toward mental health problems came at a later stage. The present study is largely concerned with opinion.

A measuring instrument was needed which was general enough to be used not only with diverse segments of the general population, but also with mental health experts. For this purpose, it seemed most appropriate that the questionnaire items should be selected from the opinions and beliefs currently held by the persons in our two sources. Opinion statements were sampled from a wide collection of materials including personal interview data, recorded interviews with the mothers of children in psychotherapy, and interviews with the relatives of mental hospital patients.[3] Additional items were sampled from mental hygiene text books and other professional publications. Samples of content from the mass media—radio, television, newspapers, and magazines—were searched for their mental health content. For example, when the "psychiatrist" in one of the radio "soap operas" advised his patient to take a long vacation, this was written down as one opinion about ways in which to treat mental health problems. In this way over 3,000 statements were gathered relating to the causes, treatment, symptoms, prognosis, incidence, and social significance of mental health problems.

The opinion statements which were collected range from the belief that "insanity" results from over-eating to the belief that psychiatrists are older than other physicians. Only those opinion statements were recorded which were expressed in a language the average person could understand. Consequently, the sample of opinions largely contained the simpler and more abbreviated opinions. Thus, there may be many opinions of a complex nature held by mental health experts which have no counterparts in our sample of opinion statements.

---

[1] The research was supported by a special grant from the National Institute of Mental Health, U. S. Public Health Service.

[2] J. C. Nunnally and C. E. Osgood are co-directors of the project. Members of the project staff during the time of the research were G. J. Suci, W. L. Taylor, and R. J. Wolff. The research assistants, T. R. Husek and Carol Tucker, gave very helpful support to the work.

[3] For lending us materials for the sampling of items, we would like to thank, among others, Angus Campbell, John Clausen, Elliot Freidson, and Shirley Starr.

The larger collection of opinions was reduced to a group of 240 by removing apparent duplicates in meaning. The 240 statements were then cast in questionnaire form. A seven-point rating scale was provided for each, ranging from "agree" to "disagree." Several examples of the opinion statements are:

> Psychiatrists try to teach patients to hold in their strong emotions.
>
> Good emotional habits can be taught to children in school as easily as spelling can.
>
> Adult problems are less important in causing emotional disorders than the individual's childhood experiences.

The questionnaire was administered to a sample of the general population in the Champaign-Urbana area. It was found that almost everyone either agreed completely or disagreed completely with 60 of the items. These were withheld from further analysis. The remaining 180 items were subjected to factor analysis. The 11 factors obtained in this way were used to describe popular beliefs about mental health problems. On the basis of the factors, a shortened version of the questionnaire was constructed containing 50 items. The 50-item form was then administered to samples of the population in Eugene, Oregon, and in Knoxville, Tennessee. This form was also used to study the opinions of experts.

The Sample of Experts

The sample consisted of 300 psychologists and psychiatrists in the United States. The 150 psychologists were randomly drawn from the diplomates in the American Psychological Association, with 75 of these being taken from the clinical section and the remaining 75 from the counseling and guidance section. The psychiatric sample consisted of 150 members of the Group for the Advancement of Psychiatry. It was thought that a sample obtained in this way could be said to represent at least a good portion of the expertness in psychology and psychiatry. Each person was mailed a copy of the questionnaire with a slightly different instruction from that used in studies of the lay population. Instead of asking each of the experts whether he agreed or disagreed with the opinion statements, he was asked whether each opinion was worthy of being advocated to the public, or used in programs of public information. Consequently, the experts used a seven-point rating scale which ranged from "support" to "repudiate."

It was explained to the experts that the middle of the continuum would be used to indicate that a particular idea should be ignored in public information efforts. Essentially, the experts were asked what messages should be conveyed to the public. In the strictest sense, the results of the study do not show what the experts believe but, more precisely, what they think the public should be told. A subsequent study of a smaller group of experts using the original "agree-disagree" questionnaire form showed such a strong correspondence with the results obtained from the "repudiate-support" form that it is reasonable to assume that the experts were responding mainly in terms of what they personally believed.

Returns from the sample were good. Of the 300 questionnaires mailed out, 201 were returned, 176 in time for our analysis. The group which was analyzed contained the responses of 86 psychologists and 90 psychiatrists.

Analysis of Data

The principal reason for getting the expert sample was to gauge the public understanding and to locate the groups in the general population whose beliefs appeared to be the most anachronistic in comparison to expert opinion. These and other findings are reported separately (1). The concern here is within the expert sample, to find the points on which they agree and disagree. First, the mean position of each of the three subgroups (two groups of psychologists and the psychiatrist group) of the expert sample on each of the questionnaire items was found. Next, standard deviations for the individual items were determined over the whole expert sample. Finally, a factor analysis was performed on 22 of the 50 questionnaire items. The particular 22 items were chosen because they showed the largest standard deviations, or disagreement, among experts.

Table 1

Rotated Orthogonal Factors

	Factor				
Item	I	II	III	IV	V
1	.24	.29	.00	−.15	.03
2	.01	.14	.35	−.23	−.18
3	.10	.27	−.02	.14	.12
4	−.22	.48	.03	−.10	−.17
5	−.03	.00	−.09	.35	−.20
6	.18	.37	.00	.22	.05
7	.57	−.03	.16	−.04	−.05
8	.40	.08	.07	.00	.43
9	.60	.03	−.12	.03	.05
10	.02	.34	−.10	−.14	.25
11	.10	.20	.08	.39	.02
12	−.10	−.12	.30	.26	.01
13	.05	.30	.17	−.09	.04
14	.22	−.23	.21	.25	.07
15	−.07	.08	.35	−.17	.29
16	.04	−.28	.10	.38	.01
17	−.23	.35	.20	.09	−.01
18	.43	−.22	.00	.10	−.10
19	.01	.33	.18	.04	−.06
20	.31	−.04	−.03	−.35	.22
21	.42	−.20	−.11	.07	−.37
22	−.19	.46	−.05	−.02	.00
23	.03	.23	.42	−.01	.00

Results

The worst fear was that the experts would show themselves no more in agreement on the questionnaire items than members of the general public. This proved not to be the case. The item variances of the experts were in all cases lower than those of the lay public and generally tended to be less than one-half[4] as large. Another way of looking at the amount of agreement among experts is in terms of the percentage of persons who disagreed with the majority decision on each item. If more than 50% of the respondents indicate that a particular item should be advocated to the public, then the percentage of persons who mark one of the "repudiate" categories, positions 1, 2, and 3, on the scale, indicates the percentage of disagreement with the majority decision. Whereas on half of the 50 items the experts show less than 10% disagreement with the majority decision, the general population shows this degree of agreement on none of the items.

The next concern was whether systematic differences would be found between psychologists and psychiatrists, or between the two groups of psychologists. The average responses (the arithmetic means) by psychologists and psychiatrists proved to be very much alike. Some small differences between professions will be shown in respect to the factor analysis work. Likewise, the average responses by the clinical and counseling diplomates were very much alike, with none of the differences approaching statistical significance.

The third consideration was whether there was sufficient disagreement on some of the items, in the context of generally high agreement, to point to schools of thought about mental health. For this purpose, the 22 items which showed the largest variance were intercorrelated and factor analyzed[5] (see Table 1). The product-moment correlation formula was used for all coefficients.

As a twenty-third variable in the correlation matrix and in the factor analysis, the profession of each subject was included. A score of 1 indicated that the subject was a psychiatrist and a score of 2 indicated that the subject was a psychologist.[6] The professional status variable was then factored in with the 22 items. The loadings of the twenty-third variable indicate the extent to which one profession is more inclined to believe a factor than is the other profession. A positive loading means that the factor is held higher by psychologists than by psychiatrists. A negative loading means that the particular factor is held higher by psychiatrists.

---

[4] Variances were averaged over the 50 items for experts and the lay population separately. The average variance for the lay population was found to be 2.3 times as large as that for the experts.

[5] The correlation matrix was factor analyzed by the principal axes method (cf. 2), sometimes called the principal components method. Communality estimates were used to fill the diagonal elements of the matrix. These were estimated by selecting the highest correlation coefficient, regardless of sign, in each column of the matrix. The five factors with the largest variance were placed through a series of single-plane rotations to achieve the orthogonal solution in Table 1. Copies of the correlation matrix, the unrotated factor matrix, and the questionnaire will be sent on request. Requests should be directed to J. C. Nunnally, Department of Psychology, University of Illinois, Urbana, Illinois. Or the material can be obtained by ordering Document No. 5486 from the American Document Institute, remitting $1.25 for microfilm or $1.25 for photocopies.

[6] When the product-moment correlation formula is applied to the relationship between one discrete and one continuous variable, the obtained coefficient is the same as the point biserial coefficient.

The most highly loaded items on the first factor are:

Factor I

Loading		*Item*
.60	[9]	Psychiatrists try to show the mental patient where his ideas are incorrect.
.57	[7]	The main job of the psychiatrist is to explain to the patient the origin of his troubles.
.43	[18]	The good psychiatrist acts like a father to his patients.
.42	[21]	Psychiatrists try to teach mental patients to hold in their strong emotions.
.40	[8]	Psychiatrists have to have a good sense of humor in order to help their patients.

Factor I concerns the psychiatrist's[7] role, and, probably, the role of psychotherapists generally. More particularly, the factor appears to concern the degree of "directiveness" that is thought to characterize the effective therapist. The items picture the effective therapists as teaching, correcting, and explaining. The factor illustrates the well-known professional controversy between those who advocate a more directive, interventionist role for the therapist, as opposed to those who advocate a role of listening and understanding. It can be seen from Table 1 that Factor I is not more characteristic of psychiatrists than psychologists. Variable 23, professional affiliation, has a loading of only .03 on the first factor. Instead, Factor I demonstrates that there is a division of opinion within both professions concerning the extent to which the therapist should direct and control the therapy situation.

Factor II

Loading		*Item*
.48	[4]	Physical rest will not prevent a mental disorder.
.46	[22]	A change of climate seldom helps an emotional disorder.
.37	[6]	Adult problems are less important in causing emotional disorders than the individual's childhood experience.
.35	[17]	Helping the mentally ill person with his financial and social problems will not cure his disorder.

The content of Factor II concerns the causes of mental disorders and the influences that make for better or worse emotional adjustment. Persons who agree strongly with these items would seem to favor an explanation of mental disorder in terms of the individual's childhood and his history of personal experience. A person who held the opposite point of view, one who would disagree with the items above, would look for the causal factors in the pressures in the individual's immediate environment. The extreme of this point of view would be that the individual's emotional adjustment is a function only of contemporary external influences. The extreme opposite point of view would be that the immediate environmental pressures are of no importance in explaining the individual's status of adjustment.

Variable 23, professional affiliation, has a small loading of .23 on Factor II. If this loading is not due to chance,[8] there is a slight tendency for psychologists to endorse Factor II more than psychiatrists. However, even if the difference between professions is real, the difference within professions is the far larger consideration.

---

[7] The term "psychiatrist" was used throughout the questionnaire to refer to an expert in the mental health field. We would have liked to use different titles, like psychiatrist, psychologist, and psychoanalyst, but this would have increased the items beyond the number that could be used in survey work. In scanning numerous personal interviews, we found that the psychiatrist is the person most often mentioned by members of the public as a person who deals with mental health problems. Also, we found that the public makes little distinction between the subprofessions in the mental health field.

[8] There are no exact error formulas that can be applied to rotated factor loadings. An analogous error term can be obtained from the standard error of the correlation coefficient. With an N of 176 in this case, the standard error would be .076, with a coefficient of .19 being significant at the .01 level. By this very approximate standard, the loading of .23 would be indicative of a real difference between the two professions.

Factor III

Loading		*Item*
.42	[23]	Professional affiliation (variable 23 in Table 1).
.35	[15]	A poor diet does not lead to feeblemindedness.
.35	[2]	Physical exhaustion does not lead to a nervous breakdown.
.30	[12]	Most of the time psychiatrists have difficulty in telling whether or not a patient's mental disorder is curable.

Factor III contains most of the common variance for the variable of professional affiliation. Rather than being a content factor, in which there should be some centralness of meaning, the factor contains an array of the items that psychologists tend to endorse more strongly than do psychiatrists. As mentioned previously, there are few items that differentiate the professions; and the differences, as evidenced in the loadings above, are small. The factor shows that psychologists are more skeptical of organic and "physical" explanations of mental disorder; i.e., diet and physical exhaustion, than are psychiatrists. The last listed item in Factor III, that concerning the psychiatrist's difficulties in prognosis, perhaps reflects more enthusiasm by some psychiatrists about their clinical prowess than would be granted by some psychologists.

Factor IV

Loading		*Item*
.39	[11]	Mental illness is one of the most damaging disorders that a person can have.
.38	[16]	Most social workers go into their field because of their own misfortunes in life.
.35	[5]	The brains of the feebleminded are smaller than those of normals.
—.35	[20]	When a person is recovering from a mental illness it is best not to discuss the treatment that he has had.

No sensible interpretation could be found for Factor IV. It is submitted to the reader for his inspection.

The loadings on Factor V are too small to be considered without additional evidence for the statistical stability of the factor. Of the common variance explained by the five factors, the percentages of variance for each were, respectively, 29.8, 28.0, 13.7, 16.6, and 12.1.

The findings in this research are restricted to the area of opinions being studied. A sample of opinions drawn more directly from what experts think and say might explain more about divisions of opinion in the mental health field. The most that can be said at this time is that there are some differences of opinions along the several lines which the factors have indicated. Other differences might be observable if new items were studied and additional factors were found.

Summary

The article concerns a questionnaire study of the opinions of psychologists and psychiatrists about mental health problems. The two professions were in agreement about most of the issues. A factor analysis was made of the combined responses for the two professions. Two of the obtained factors demonstrate schools of thought about mental health problems. A third factor contains a number of items which serve to differentiate the opinions of psychologists and psychiatrists.

Received June 3, 1957.

References

1. Nunnally, Jum. The communication of mental health information: A comparison of the opinions of experts and the public with mass media presentations. *Behav. Sci.*, 1957, 2, 222–230.
2. Thurstone, L. L. *Multiple factor analysis.* Chicago: Univer. of Chicago Press, 1947.

Journal of Consulting Psychology
Vol. 22, No. 3, 1958

Some Social-Psychological Characteristics of a Child Guidance Clinic Caseload[1]

Jack L. Roach[2]
Guidance Center of Buffalo

Orville Gurrslin
Youth Board of Buffalo

and Raymond G. Hunt
University of Buffalo

A number of recent studies have indicated that social-psychological variables play a significant role in determining the utilization of mental health facilities by members of the community (1, 7, 9, 10). While such a phenomenon is generally in accord with certain theoretical expectations (4), it is difficult to draw generalized conclusions from the available literature. By and large, studies in this area have dealt with tendencies in specific geographic areas and, usually, specific mental health facilities. As a result, the data obtained may be seriously biased as a function of idiosyncratic features of the particular area or agency involved. Thus the degree to which sources of data are representative of the universe to which this data becomes generalized is an especially critical issue in studies of this type.

The sampling problem can perhaps be resolved in a number of ways, probably the most efficient of which is replication of such studies with independent samples in different geographical areas. The pooled findings of these studies then constitute descriptive steps permitting analytic or systematic generalization concerning the sociology and social psychology of mental health and institutional structures in that area.

The present investigation then, while not a direct replication of any particular study, is an attempt to investigate the relationship between a number of social-psychological characteristics found to be of significance by other investigators and the caseload of a community child guidance clinic.

Method

Case record information was obtained from 246 cases served in some capacity at the Guidance Center of Buffalo during 1954 and 1955. The Guidance Center is a community supported child guidance clinic with services available, through an open referral system, to the residents of Erie County, New York. Fees are adjusted according to ability to pay, there being no reason for a family to be refused service because of low income.

The study includes the entire child caseload active at some time during the period covered with the exception of cases wherein the child was not residing with his own parent(s), either natural or adoptive. Thus, institutionalized or foster-home children were excluded from the study.

Data pertaining to several social-psychological characteristics of the families receiving clinic services were extracted primarily from the application form filled out by the parents at intake. As generally happens in studies of this sort, case records do not always yield complete information about every case. As a result, in each descriptive category employed there was a varying number of cases which could not be ordered exactly. After careful consideration, however, the authors deem it unlikely that the findings reported below would be seriously altered were all of the unknown cases known. Comparative census data, where appropriate, were drawn from the Statistical Abstract of the United States (11), and from records of the Council of Social Agencies of Buffalo and Erie County.

---

[1] The authors wish to thank A. V. Cutter, director, and the staff of the Guidance Center of Buffalo whose cooperation made this study possible.

[2] Now with the New York State Department of Social Welfare.

Seven characteristics presumed to be of significance were selected for analysis. These were: (*a*) sex of child, (*b*) family size, (*c*) occupation of major wage earner, (*d*) area of residence, (*e*) race, (*f*) religious affiliation, and (*g*) symptom picture.

The purpose of the present study was entirely descriptive (empirical). No attempt was made, at this time, to establish relationships between any of the variables, though a number of interrelations are likely as suggested below.

Results

It was found that of the 246 cases included in this study, 72% were male. This finding is consistent with data reported by several other investigators (1, 7), thus indicating a tendency toward disproportionate representation of males in child guidance clinic patient populations.

Table 1

Occupational Distribution of Major Wage Earner in Clinic Families

($N = 241$)

Occupational group	Source (in per cent)	
	Clinic[a]	Census[b]
Professional and technical	21	11
Proprietors, managers, officials	20	12
White collar, clerical	19	24
Skilled workers	19	15
Semiskilled workers	11	31
Unskilled workers	10	7
Total	100	100

Note.—$\chi^2\% = 24.77$, $df = 5$, $p = <.001$.

[a] Unknowns (7% of total) were distributed among categories according to observed distribution of known cases.

[b] Census figures refer to New York State (1950).

Table 2

Quartile Ratings of Desirability of Residential Area

($N = 196$[a])

Quartile ranks	Source (in per cent)	
	Observed (clinic)	Expected
I	52	25
II	11	25
III	24	25
IV	13	25
Total	100	100

Note.—$\chi^2\% = 57.82$; $df = 3$; $p = <.001$.

[a] The bulk of unknown cases (20%) omitted from the analysis came from tract areas not rated but primarily desirable suburban areas.

A similar lack of representativeness was observed with respect to the racial characteristics of the clinic population—only 1% of which was nonwhite. In both instances, the significance (statistical) of the data is obvious.

Table 1 reports the results of a comparison between the occupational distribution of the major wage earner in the clinic population and the distribution of such occupations in the general population. The classification employed was a modification of occupational classifications frequently reported in the sociological literature (2). Farmers and unemployed workers were excluded from the analysis owing to the very low frequency of each in the clinic population (two and three cases, respectively). Moreover, it would be expected that few farmers would be represented owing to the urban character of the area served by the clinic. Hence, to avoid the introduction of spurious differences, these groups were omitted from analysis.

Table 1 indicates that the agency is serving all occupational groups within the general population, but with a considerable overrepresentation of the "upper" occupational categories.

The data reported in Table 2 were obtained by rating the area of residence of each clinic family along a dimension of desirability, according to a quartile scale developed by the Council of Social Agencies of Buffalo and Erie County (3). This scale is based upon a

Table 3

Religious Affiliation of Clinic Families

($N = 246$)

Religion	Source (in per cent)	
	Clinic[a]	Census[b]
Protestant	57	40
Catholic	32	57
Jewish	11	3
Total	100	100

Note.—$\chi^2\% = 33.50$; $df = 2$; $p = <.001$.

[a] Unknowns (20%) were distributed according to observed frequencies of known cases.

[b] Census figures refer to metropolitan Buffalo (1955).

variety of objective characteristics of census tracts in Buffalo and vicinity (e.g., percentage of dilapidated dwellings). It should be noted that a rating of I corresponds to "most desirable" and a rating of IV to "least desirable."

It can be seen from Table 2 that a heavy proportion of clinic families come from the most desirable residential areas. It should be borne in mind that the variables summarized in Tables 1 and 2 involve dimensions widely used as indices of social-class position.

Data pertaining to the religious composition of the clinic caseload are presented in Table 3. Both the present study and Anderson and Dean (1) agree generally in finding that Protestants comprise the largest single clinic group. Both this and the Jewish group are overrepresented in the present clinic population, while Catholics are considerably underrepresented.

Table 4

Number of Children in Clinic Families

($N = 246$)

Number of children	Source (in per cent)	
	Clinic[a]	Census[b]
1	15	35
2	41	33
3	20	18
4 and more	24	14
Total	100	100

Note.—$\chi^2\% = 20.86$; $df = 3$; $p = <.001$.

[a] Unknowns (3%) were distributed according to observed frequencies of known cases.

[b] Census figures refer to New York State (1955).

In Table 4 are presented data relating to the number of children in clinic families. It is clear that the obtained difference is accounted for mainly by the underrepresentation of single child families in the clinic population. It is of considerable interest that only 15% of the case population is made up of "only" children—a finding contrary perhaps to expectancies derived from much child development literature (5).

Table 5 represents a classification of the symptomatic picture characterizing the child at intake. The classification was developed by the authors. Cases were ordered to categories on the basis of dominant symptom patterns. When no particular type of symptom was dominant, the case was classified as "mixed," hence this category is somewhat more heterogeneous than the others. As far as possible, ratings were based upon descriptions of the child's behavior provided by the parent(s). An effort was made to sift out interpretive statements originating with clinic personnel. It was hoped thereby to lend greater objectivity to the data.

Evident in Table 5 is the marked incidence of children with aggressive behavior as the predominant symptom. This seems generally consistent with the findings of Stevens (10) and of Anderson and Dean (1). Thus, aggressive and acting-out behavior apparently com-

Table 5

Classification of Dominant Symptomatic Behavior at Intake

($N = 246$)

Dominant symptoms	f	%
Aggressive	86	35
Nonaggressive[a]	69	28
Somatic complaints[b]	23	9
Mentally retarded[c]	35	14
Mixed	25	10
Unknown	8	4
Total	246	100

[a] For example, withdrawal, daydreaming, "silly behavior," excluding somatic functions when dominant complaint.

[b] For example, asthma, pains in stomach, when dominant complaint.

[c] Cases classified here when dominant complaint was retardation, suspicion thereof, or poor academic performance.

prises the largest single basis for clinic referral, perhaps because such behavior is most disturbing to various cultural agents.

Finally, it is perhaps worth noting that 25% of the patient population included in this study had had some kind of prior social agency contact as reflected by registration with the Social Service Exchange.

Discussion and Conclusions

A general conclusion which emerges from the findings reported here is that the clinic from which the data was drawn is not serving equally all segments of the population. It is clear that the clientele served represents, primarily, a white, Protestant group of middle to high socioeconomic status. Such also is the conclusion drawn by other similar studies.

Further, it is quite conceivable that most of the observed instances of unrepresentativeness of clinic populations are a direct function of the middle-class character of such populations. If, for example, the population served is largely middle class, it is not at all surprising that nonwhites will form a very small percentage of clinic clientele. Likewise, the convergence between Protestant and general middle-class values pointed out by many writers (Davis, 4; Merton, 8; Williams, 12), forms one plausible explanation for the observed distribution of religious groups in the clinic population.

Further, the fact that a large percentage of Catholics is included in the lower social classes renders understandable the underrepresentation of this group in the population of urban child guidance clinics, although other factors might also be operative with regard to this group. The same considerations would hold, in reverse, as an explanation for the observed overrepresentation of Jews.

Similarly, this "middle-class phenomenon" might be of considerable significance with respect to the preponderance of male children served (Green, 6), and also with regard to the dominant type of symptom patterns observed. Regarding this latter, it seems likely that instances of acting-out behavior will be a source of greater concern among middle-class than among working-class families.

It thus becomes apparent from a growing body of literature similar to the present study, that cultural mental health facilities are being utilized primarily by middle- and upper-class members of society. The independent variables determining this phenomenon are as yet, however, not clearly understood and obviously call for empirical elucidation.

Received June 24, 1957.

References

1. Anderson, F. N., & Dean, H. C. Some aspects of child guidance clinic intake policy and practice. U. S. Dept. of Hlth, Educ. & Welf., *Publ. Hlth Monogr.*, 1956, No. 42.
2. Bendix, R., & Lipset, S. *Class, status and power*. Glencoe, Ill.: Free Press, 1953.
3. Council of Social Agencies of Buffalo and Erie County, N. Y. *Buffalo self study of recreation, informal education and group work services*. Buffalo, N. Y.: Author, 1956.
4. Davis, K. Mental hygiene and the class structure. In P. Mullahy (Ed.), *A study of interpersonal relations*. New York: Hermitage House, 1949. Pp. 364–386.
5. Epperson, J. A. Some basic principles of direct work with children. *Child Welfare,* 1952, **21**, 3–8.
6. Green, A. The middle-class male child and neurosis. In L. Wilson & W. Kolb (Eds.), *Sociological analysis*. New York: Harcourt, Brace, 1949. Pp. 236–248.
7. Hollingshead, A., & Redlich, F. Social stratification and psychiatric disorders. *Amer. sociol. Rev.*, 1953, **18**, 163–169.
8. Merton, R. K. *Social theory and social structure*. Glencoe, Ill.: Free Press, 1949.
9. Michigan Department of Mental Health. Differential utilization of the facilities of a Michigan child guidance cilnic. *Mich. Dept. ment. Hlth res. Rep.*, Mar. 1955, No. 17.
10. Stevens, S. An ecological study of a child guidance clinic intake. *Smith Coll. Stud. soc. Wk.*, 1954, **25**, 73–84.
11. U. S. Dept. of Commerce, Bureau of the Census. *Statistical abstract of the United States*. Washington: U. S. Govt. Printing Office, 1956.
12. Williams, R. *American society*. New York: Knopf, 1951.

Journal of Consulting Psychology
Vol. 22, No. 3, 1958

The Weighted *Q* Sort: A Procedure for Quantitatively Estimating Emotional Disturbance and Personality Change[1]

Marvin Reznikoff and Laura C. Toomey
Institute of Living, Hartford, Connecticut

While the usefulness of the *Q* sort has been clearly demonstrated in expressing individual personality profiles in quantitative terms, the technique is not equally effective for quantitative representations of personality change or degree of emotional maladjustment. In a typical application of this forced choice rating method, *Q*-sort personality evaluations are done on patients before and after therapy and the sorts compared. Coefficients of correlation indicate if personality changes, per se, have occurred, but do not reflect quantitatively the direction and magnitude of such changes; neither do *Q* sorts used in this manner provide direct estimates of levels of adjustment which can be contrasted. Only Dymond (3) has published a method for quantitatively evaluating adjustment using an "external criterion." Dichotomizing her *Q*-array statements into "adjustment" and "maladjustment" categories, she arrives at an adjustment score by counting the adjustment items characteristic of the patient and the maladjustment items not characteristic of him.

As one aspect of a large project involving the evaluation of psychological changes associated with psychiatric treatment, a modification of the *Q* sort was developed which yields, along with the usual personality picture, a single composite score revealing degree of emotional disturbance. Additionally, this "adjustment index" readily lends itself to an over-all quantitative determination of personality change. In contrast to Dymond's method of broadly classifying items as indicative of either adjustment or maladjustment, this procedure places items at appropriate points on a graduated adjustment-maladjustment scale. The modified technique designated as "the weighted *Q* sort" is described in this paper.

Development of the Method

As a first step, a group of clinical formulations were selected from Beck's psychodynamic outline of schizophrenia (1) and modified to some extent. Since Beck's conceptual schema is based solely on Rorschach personality correlates, pertinent clinical formulations were also included which could be determined more readily from other projective tests, as well as from information elicited during therapeutic sessions. Finally, some statements characteristic of personality in general were added, in order to more fully represent the normative end of the continuum, and provide a schema applicable to nonschizophrenic as well as schizophrenic subjects. All 44 clinical statements were then typed on individual 3″ × 5″ cards and sorted into a seven-category forced normal distribution by three experienced clinicians. These judges, working independently, rated these items on a continuum from most indicative of severe emotional disturbance (Category 7) to least indicative of severe emotional disturbance (Category 1). The specific sorting instructions provided were:

> Please sort these 44 clinical formulations into seven categories. Category 7 is to represent those clinical formulations most indicative of a serious psychological disturbance. The clinical formulations in the remaining six categories are to be thought of as progressively less indicative of a serious mental disturb-

---

[1] This investigation was supported by the Office of Naval Research under Contract No. NONR (00) (01).

ance with those items in Category 1 considered to be least pathological in their implication. The number of cards to be placed in each of the categories is as follows:

Category	7	6	5	4	3	2	1
Number of cards	3	6	8	10	8	6	3

The correlations between the ratings of the three clinicians taken two at a time were .84, .75, and .73, indicating a statistically significant degree of agreement ($p < .001$) in evaluating the relative pathological implications of the clinical formulations. Using the totals of the three independent ratings for each of the 44 items distributed in an approximate normal curve, a weight or "pathology score" was assigned to every clinical formulation. The three clinical formulations having the highest total ratings were each given a weight of "7," designating them as most pathological; the six having the next highest totals a weight of "6" and so on, with decreasing weights representing diminishing pathology. The items and their pathology scores are:

Pathology score of 7 (3 items)
Flight of ideas
Delusional thought
Thought processes are regressive, archaic, primitive or magical

Pathology score of 6 (6 items)
Inappropriate affect
Metaphysical or philosophical ponderings of a vague unproductive nature
Grandiose
Withdraws into autistic fantasy
Confused, illogical thinking
Erotic gratifications regressive

Pathology score of 5 (8 items)
Prominent mood swings
Disregards social canons
Intrusion of irrelevant material in thinking
Projection an important defense
Suspicious and wary
Inadequate reality testing
Free anxiety; intense and sudden in appearance with controls failing
Maladaptive externalized aggression

Pathology score of 4 (10 items)
Socially withdrawn
Primarily identified with opposite sex
Aggressive, destructive preoccupations
Interest is dominated by one or two themes to the exclusion of all others
Impulsive, unstable in reaction to usual life stimuli
Intellectual functioning significantly below potential
Denial an important defense
Sexual preoccupation
Undoing an important defense
Depression

Pathology score of 3 (8 items)
Blocking
Generalized feelings of inadequacy and ineffectuality
Isolation an important defense
Guilt feelings
Anxiety provoking conflicts expressed in conversion symptoms or psychosomatic illness
Infantile and egocentric
Anxiety persistent and pervasive
Obsessive-compulsive behavior

Pathology score of 2 (6 items)
Repression an important defense
Displacement an important defense
Unrealistic self-concept and level of aspiration
Reaction formation an important defense
Passive and submissive
Characterized by rigidity

Pathology score of 1 (3 items)
Rationalization an important defense
Overconventional
Limited insight

A convenient arrangement is to note the appropriate pathology score on the back of each clinical formulation card. As employed in personality evaluation, this weighted Q array is sorted in the ordinary manner. The clinician sorts each of the clinical formulations into one of seven categories, depending on how characteristic the description is of the patient. The number of items to be placed in each category is predesignated, so that an approximate normal distribution, identical with the one used for the pathology sort, is obtained. Items in Category 7 are those judged most applicable to the patient, while the formulations in Category 1 are considered entirely inappropriate descriptions.

Denoting the category in which a particular clinical formulation is placed in a personality assessment as its "applicability rating," a composite adjustment index can be computed for a patient by simply multiplying each applicability rating by the item's pathology score and summing the 44 products. If the clinical formulations indicative of the most severe pathology are always rated as most characteristic of the patient, then a maximum total will be obtained, reflecting the greatest degree of disturbance. On the other hand, if the items judged as most pathological are consistently scored as least characteristic of the patient, the total is minimal, indicating relatively insignificant emotional disturbance. The highest possible total for the 44 clinical formulations using this procedure is 722, while the lowest total is 586. A range of 0 to 236 can thus be established for the adjustment index by subtracting the constant 586.

Reliability

As a check on the reliability of the adjustment index, weighted *Q*-sort personality evaluations done independently on the same patients by each author were contrasted. These sorts were based on the test protocols of 25 patients not tested by either author and selected at random from the files. In general, the battery of tests available on each patient consisted of the Wechsler, Rorschach, Thematic Apperception Test, Draw-a-Person Test, and the Sentence Completion Test. The correlation between the two series of adjustment indices, calculated from the two independent *Q* sorts for each patient, was .79, which is significant at .001 level, and indicates that this procedure can be used in consistent fashion.

Validity

To determine the validity of this procedure, adjustment indices from the two *Q*-sort ratings for each of the 25 patients were averaged, and the mean ratings were compared with estimates of adjustment obtained from each patient's therapist. Each of the psychiatrists involved was asked to rate the patient as he was at the time of the psychological testing, using The Institute of Living Clinical Rating Scale for Attending Physician (5). This scale consists of 12 items, each to be rated on a five-point scale of increasing maladjustment. The ratings may be summed to give an over-all estimate of adjustment, with a maximum total score of 60 corresponding to maximum maladjustment. The total scores from the therapist rating were found to be correlated with the mean *Q*-sort indices .57, a result which is significant at the .01 level. Correcting for attenuation increases this correlation coefficient to .66.

Discussion

This modified procedure has a number of advantages compared with the conventional correlation approach. First, the computation of the adjustment index is somewhat less arduous than the derivation of a correlation, although the latter computation can be simplified by use of the "method of differences" (2, 4). Also, the adjustment index escapes certain disadvantages inherent in the *r:* the potentiality for being mistaken for a percentage, and the complexity of interpretation of a measure that may vary in significance.

But the most important advantage of the adjustment index approach lies in its ability to compare *Q* sorts from *groups* of patients, and in its amenability to further relatively simple statistical manipulation. Returning to the example of contrasting "unweighted" *Q*-sorts personality evaluations done on a number of patients prior to and after therapy, the method of comparison is a correlation, calculated separately for each patient. The researcher is thus left with a distribution of *r*'s which may be exceedingly difficult to interpret. When the weighted *Q* sort is used, however, pre- and posttherapy adjustment indices are computed for every patient in the manner described previously. A statistical test is performed next to determine if the mean difference between the two series of adjustment indices is significant. This procedure then gives a definitive picture of the scope and direction of changes associated with therapy.

The advantage of the index is even clearer in the case where two groups of patients are to be compared before and after therapy. The investigator using correlations would be faced with the task of comparing two series of *r*'s, the before-after therapy correlations for the two groups. The *r*'s cannot be compared directly, but must be converted to *z* values; whereas two series of adjustment scores differences could be readily compared by conventional tests of significance. Furthermore, the sign of the index difference immediately indicates the direction of change, while direction cannot be discerned from the *r* comparison.

The weighted *Q* sort also readily lends itself to a comparative analysis of the ratings of psychiatrist and psychologist judges, in that the 44 clinical formulations can be rated from the results of psychological tests or therapeutic interviews. Of particular interest would be the use of weighted *Q* sorts to determine the degree of agreement between psychologists and psychiatrists evaluating the general adjustment of the same group of patients.

The authors' experience with the weighted *Q* sort in its present form suggested that augmenting the 44 clinical formulations with additional items would enhance the sensitivity of the technique.

Summary

A method was described for modifying the *Q* sort so that it could be utilized to estimate over-all degree of emotional disturbance and also to determine extent of personality change. Briefly, the procedure involved weighting each of 44 *Q*-array clinical formulations according to how indicative the specific item was of serious pathology. The statements can then be sorted, in the usual manner, into categories arranged on continuum: personality descriptions judged most characteristic of the patient to be placed at one end of the continuum, while those rated as least characteristic at the other extreme. From the relative applicability of the various clinical formulations to the particular patient and the assigned weights of the items, an individual "adjustment index" is determined. Reliability and validity data with reference to this procedure are presented and the advantages of using the weighted *Q* sort's adjustment indices in group comparisons are discussed.

Received January 9, 1958.
Early Publication.

References

1. Beck, S. J. *The six schizophrenias.* New York: American Orthopsychiatric Association, 1954.
2. Cohen, J. An aid in the computation of correlations based on Q sorts. *Psychol. Bull.*, 1957, **54**, 138–139.
3. Dymond, Rosalind F. Adjustment changes over therapy from self-sorts. In C. R. Rogers & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954. Pp. 76–84.
4. Hilden, A. H. *Manual for Q-sort and random sets of personal concepts.* St. Louis, Mo.: Author, 1954.
5. Reznikoff, M., & Zeller, W. W. A procedure for evaluating the status of schizophrenic patients. *J. clin. exp. Psychopath. & Quart. Rev. Psychiat. Neurol.*, 1957, **18**, 367–371.

Journal of Consulting Psychology
Vol. 22, No. 3, 1958

Adjustment and Prediction of Others' Behavior

June Elizabeth Chance
University of North Carolina

Among the myriad problems in the area of interpersonal perception are those concerning the characteristics of the individual perceiver which affect the adequacy of his perceptions. Implicit in many conceptions of maladjustive behavior is the notion that the poorly adjusted individual is less able to perceive clearly and hence to predict the responses of others. However, empirical evidence regarding the relationship between personal adjustment and ability to predict is equivocal (5). A recent study by Bieri, Blacharsky, and Reid (1) hypothesized a negative relationship between degree of maladjustment and accuracy of prediction, as well as a positive relationship between degree of maladjustment and tendency to engage in projection (assuming similarity between self and others). Neither hypothesis was supported by the findings; in fact, the obtained relationship between degree of maladjustment and projection was in the direction opposite from that predicted. On the basis of their study, these investigators concluded that "better adjusted *S*s tend to predict most accurately on the basis of *similarities* between themselves and others, while maladjusted *S*s emphasize to a greater degree than adjusted *S*s accurate prediction of differences between themselves and others" (1, p. 353).

Some methodological aspects of this study suggest, however, an alternative explanation of the findings as has been recently pointed out by Erikson (3). The choice of the Taylor Manifest Anxiety Scale as the set of responses to be predicted leaves open the possibility that *S*s in making their predictions could have been responding to a stereotype of how a college student responds. More specifically, if most *S*s were aware for most items of the MAS that certain responses were more "acceptable" and probably more likely to occur in a college population, the results obtained by Bieri, Blacharsky, and Reid would automatically follow. An *S* who himself had given maladjusted responses would "predict" the responses of his target figures to be mostly different from his own and consequently a high proportion of his accurate predictions would be perceived differences. Conversely, the well-adjusted *S* responding in the more acceptable and likely manner for himself and also responding in this manner for his targets would obtain most of his accuracy in perceived similarities. Further, in this study *S*s chose members of their class to predict, and since differential acquaintance with fellow class members might be related to the variable of adjustment, this procedure could easily confound the results of the study.

The present study is an attempt to test the hypothesis that there is a negative relationship between personal adjustment and adequacy of performance in predicting the behavior of others. The method used somewhat obviates the possibility that *S*s can predict on the basis of a stereotype and also the difficulties which may ensue in interpreting results where *S*s select acquaintances to predict.

Method

Fifty-two *S*s from two sections of an elementary psychology course at the University of North Carolina made predictions for their respective section instructors. Both groups were administered the Rotter Incomplete Sentences Blank (4) and a modified group form of the Minnesota Multiphasic Personality Inventory, which included the Welsh *A*[1] and *R* Scales (7). At a later time, after approximately 35 hours of class meetings, the prediction procedure was administered. *S*s were asked first to respond to the prediction schedule for themselves and immediately after to predict how their respective instructor had responded to the same items. All experimental procedures were administered by an experimenter[2] who had nothing to do with the teaching of the course, and *S*s were assured that their instructors would not see the data until after the end of the term.

The criterion instrument on which the predictions were made was constructed by selecting 40 items from the Edwards Personal Preference Inventory (2). The inventory is an attempt to measure psychological needs by means of a forced choice procedure. Each item consists of a pair of statements from which the individual is to select the one most descriptive of him. Each pair of statements is equated for social acceptability; that is, each member of a pair is about equally likely to be chosen in a college population. The 40 item pairs used for the prediction schedule included 10 pairs each from the scales measuring needs for achievement, affiliation, aggression, and autonomy. The choice of items was a priori, but with two objectives—to provide a variety of behaviors to be predicted and behaviors of a sort that *S*s would be likely to possess some relevant observations from which to make their inferences.

The target figures were the instructors of the two sections: one instructor was male, one female; each section predicted its own instructor. The sections met at the same hour and one half of the students enrolled in the course had been arbitrarily assigned to each. Both instructors had taken the Edwards' PPI out of curiosity about six months before the present study was suggested.

Scores for predictive behavior were derived in the same manner as in the Bieri, Blacharsky, and Reid study. Predictive accuracy scores were obtained by totalling the number of agreements between *S*'s predictions and the target's responses. Assimilative projection scores were obtained by totalling the number of *S*'s predictions (accurate and inaccurate) which coincided with his own responses. The latter score was separated into two components: accurate projection, i.e., accurately perceived similarities; and inaccurate projection. A perceived difference score was obtained by totalling the number of disagreements between *S*'s predictions and his own responses. This score was also separated into two parts: accurately and inaccurately perceived differences. In addition, the number of real similarities and the number of real differences between each *S* and the target were totaled.

Results

The Welsh scales and ISB yielded distributions of scores very similar to those reported by the authors for other college populations. Two judges independently scored 28 of the 52 ISB protocols; the obtained interscorer reliability was .89. The correlation between the Welsh *A* scale and the ISB was .38; between the *R* scale and the ISB, .30; between *A* plus *R* and the ISB, .35. All of these correlations among the various measures of adjustment are significant beyond the .05 level. (The correlation of .04 between the *A* and *R* scales for this group is not significantly different from zero.)

The prediction task given the *S*s was a difficult one. The mean predictive accuracy score for all *S*s was 22.39 (*SD* 1.99). Since a predictive accuracy score of 20 could presumably be obtained by chance, a test of the significance of the difference between the obtained mean and 20 was performed (6, p. 147). The value of t equalled 7.68 ($p < .001$, df 51).

The product-moment correlations between the various prediction scores and the measures of adjustment are presented in Table 1. Fisher's z transformation was used to test the

---

[1] Welsh suggests that the *A* scale measures an aspect of maladjustment characterized by anxiety and dysphoria. Mr. Carl Cochran, in a doctoral dissertation in progress, found a correlation between the MAS and the Welsh *A* scale of + .81 for a group of 514 college students. The *R* scale was included to get at another aspect of adjustment—tendency toward denial and repression.

[2] The author wishes to thank Howard Leventhal for his assistance in the collection of these data.

Table 1

Relationships Between Measures of Adjustment and Prediction Scores ($N = 52$)

Measure	ISB	*A* scale	*R* scale
Predictive accuracy	−.26*	.02	−.14
Assimilative projection	.05	−.28**	.02
Accurate projection	.00	−.18	.05
Inaccurate projection	.11	−.26*	.21
Perceived differences	−.05	.28**	−.02
Accurately perceived differences	−.37***	.16	−.02
Inaccurately perceived differences	.16	.19	−.02

* Indicates $p < .10$.
** Indicates $p < .05$.
*** Indicates $p < .01$.

significance of the differences between the obtained *r*'s and zero. For all the measures of adjustment a high score is indicative of maladjustment.

The correlation between predictive accuracy and ISB scores although in the expected direction did not quite reach statistical significance ($.10 > p > .05$). Neither *A* nor *R* correlated significantly with predictive accuracy.

While ISB scores do not correlate with tendencies to assume similarity, either accurately or inaccurately, scores on the *A* scale do seem to be negatively related to inaccurate projection (assumption of unwarranted similarity), meaning that the higher the *A* score the less likely the person to assume unwarranted similarity between himself and the person predicted for. Although the positive correlation of *R* with inaccurate projection does not quite reach the .10 level, it suggests that the independent aspects of adjustment measured by the *A* and *R* scales are related to tendencies "to see similarities where they are not" in quite different ways. Also inspection of the relationships between the three projection scores and the *A* and *R* scales suggests that while the negative relationship between *A* and inaccurate projection could well be largely the result of some general response set to assume dissimilarity, the positive relationship between *R* scores and inaccurate projection might be independent of such a set.

ISB scores are not significantly related to over-all perception of differences between self and the target predicted; however, they are negatively related ($p < .01$) to accurate perception of differences, suggesting that maladjusted *S*s predict differences less well than do better adjusted *S*s. In contrast, scores on the *A* scale are positively related to over-all perception of differences, although neither the accurate nor inaccurate components of these predicted differences correlate significantly with scores on the *A* scale.

Real differences between *S* and the predicted target correlated with predictive accuracy − .07, which is not significant. Real differences correlated + .02 with scores on the *A* scale and − .16 with scores on the ISB. Accurately predicted differences correlated with real differences + .64. When a partial correlation between accurately predicted differences and ISB scores holding real differences constant is computed the value is − .36.

Discussion

The above findings do not support Bieri, Blacharsky, and Reid's conclusions, i.e., that better adjusted *S*s are more likely to predict similarities accurately, while more poorly adjusted *S*s are more likely to predict differences accurately. The obtained results do suggest that there may be a general response set to assume similarity or difference when making predictions related to the dimension of adjustment measured by the *A* scale, but not related to the dimensions measured by either the ISB or the *R* scale. Maladjustment as measured by the *A* scale seems to be related to a tendency to believe that others are different from one's self; hence, in a study where the targets predicted are truly quite different from the *S*s, spurious positive correlations between maladjustment and predictive accuracy could be obtained.

The hypothesis of a negative relationship between adjustment and adequacy of predictive behavior is neither clearly refuted nor confirmed. The findings demonstrate a trend toward a negative relationship between ISB scores and over-all accuracy of prediction in the kind of experimental situation described here; this negative relationship is most apparent in the differential ability among individuals of different levels of adjustment to predict differences between self and others accurately.

Generalizations of these findings are limited by the fact that the three measures of adjustment used, while positively correlated among themselves, are related to the predictive accuracy scores in different ways. Statements about the relationship of adjustment to prediction behavior are not meaningful apart from the particular criteria of adjustment employed. Although the study as executed has the virtue that all *S*s had rather extended and fairly uniform acquaintance with the targets predicted, it must be pointed out that most other studies in this area have selected target figures who were either *S*'s peers or who could readily be identified by *S* as peers. To the degree that *S*s might be expected to operate with a different set of assumptions about figures approximately 10 years older than they and with Ph.D.'s in psychology, the results of the present study may not be comparable to those of other studies.

Perhaps better answers in this exceedingly complex area could be obtained if the questions asked were framed differently. That is to say, if questions were asked not in terms of highly abstracted variables like adjustment-maladjustment, but rather in terms of what aspects of adjustment or what particular and characteristic modes of dealing with problems might be related to adequacy of interpersonal perceptions. Further studies could also profit from more specification of what is predicted. One might hypothesize that given modes of adjustment are differentially related to ability to predict certain classes of behavior on the part of others. For example, individuals who have difficulty in coping with their own aggressive propensities might be expected to be less accurate in predicting aggressive responses on the part of others, but relatively more accurate in predicting others' dependent responses.

Summary

Fifty-two college students were asked to respond to a 40-item, forced choice questionnaire and then to predict their respective instructor's responses to the same questionnaire. On a previous occasion, three paper-pencil measures of adjustment had been obtained for these *S*s—the Rotter Incomplete Sentences Blank, and the Welsh *A* and *R* scales (MMPI). Scores for over-all accuracy of prediction, various components of accuracy, and tendencies to assume similarities or differences between self and others in making predictions were correlated with the measures of adjustment. Results obtained indicate a trend for adjustment as measured by the ISB to be negatively related to adequacy of prediction. Scores on the *A* scale were negatively related to tendency to assume similarity in making predictions.

Received June 3, 1957.

References

1. Bieri, J., Blacharsky, E., & Reid, J. W. Predictive behavior and personal adjustment. *J. consult. Psychol.*, 1955, **19**, 351–356.
2. Edwards, A. L. *Manual: Edwards Personal Preference Schedule.* New York: Psychological Corp., 1954.
3. Erikson, C. W. Personality. *Annu. Rev. Psychol.*, 1957, **8**, 185–210.
4. Rotter, J. B., & Rafferty, Janet E. *Manual for the Rotter Incomplete Sentences Blank, College Form.* New York: Psychological Corp., 1950.
5. Taft, R. The ability to judge people. *Psychol. Bull.*, 1955, **52**, 1–23.
6. Walker, H. M., & Lev, J. *Statistical inference.* New York: Holt, 1953.
7. Welsh, G. S. Factor dimensions A and R. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine.* Minneapolis: Univer. of Minnesota Press, 1956.

Journal of Consulting Psychology
Vol. 22, No. 3, 1958

Nonprojective Determinants of Perception on the TAT[1]

Bernard I. Murstein
Louisiana State University

Projective techniques are coming under ever closer scrutiny with regard to the meaning of the material elicited by these techniques. Since the various techniques differ in their stimulus properties (clay, ink blots, dolls, scenes of people, etc.), it would seem of importance to investigate the nonprojective determinants of perception as well as the more personal aspects, in order to interpret the content in a more meaningful manner.

Eron in earlier studies (1, 2) showed that the emotional tone of the stories told to the TAT cards was more often negative (unhappy, sad, hostility, death) than positive (pleasant, friendly, contented).

It should follow, therefore, that a positive correlation ought to exist between the emotional tone and the positive-negative stimulus value of the cards such that negative scenes tend to elicit negatively toned stories.

In addition, clinical experience with the TAT cards has suggested to the author that in those cards which are relatively unstructured or ambiguous with regard to their stimulus properties, positively-toned stories were often elicited. Conversely, clearly structured or unambiguous cards often elicited negatively-toned stories. The last observation does not necessarily imply any innate relationship between structure and tone, but rather suggests that in their selection of cards in constructing their tests, Murray and Morgan chose many highly structured cards whose stimulus properties also were quite negative with regard to the emotions expressed.

The following hypotheses were therefore formulated:

1. The emotional tone of stories told in response to the TAT cards is positively related to the pleasantness-unpleasantness of the TAT cards. Those cards pleasant in appearance yield pleasantly toned stories.
2. The emotional tone of stories is also a function of the clarity of the emotions expressed on the TAT cards. Those cards most clearly structured (unambiguous) will yield negatively toned stories while unstructured cards yield more positive themes.
3. While some overlap exists between the dimensions of pleasantness-unpleasantness, and ambiguous-unambiguous, both dimensions show sufficiently low intercorrelation to be considered separately as factors determining the tone of the stories told.

Procedure

Twelve normal female subjects (*S*s) (8 students, 4 secretaries), ranging in age from 18 to 39, with a median age of 23, were given, in a randomized fashion, the same 20 TAT cards that Eron (2) used in his earlier study with women (1, 2, 3 GF, 4, 5, 6 GF, 7 GF, 8 GF, 9 GF, 10, 11, 12 F, 13 MF, 14, 15, 16, 17 GF, 18 GF, 19, 20). The age range and median were roughly equivalent to that of Eron's group.

The emotional tone ratings quantified by Eron on a five-point scale ranging from 2 to −2 according to their pleasantness-unpleasantness were totaled up for each card from his study, omitting rejects. These sums were then converted into ranks so that the most unpleasantly toned card was ranked 1, while the most pleasant card was ranked 20.

The 20 cards were given individually by four different experimenters to the 12 *S*s, with the author administering the cards to six *S*s, while the others administered the cards to two *S*s each.[2] Essentially, the following instructions were given:

---

[1] Paper read at the meeting of the American Psychological Association, New York, 1957.

> I want you to rank each card according to how clear the card is with regard to the emotions expressed in it. It does not matter whether the card is happy, sad, bored, friendly, etc. That card in which the emotion expressed is most easily determined you will rank as 1. That card which is next clearest in terms of some definite emotion expressed, rank as 2, and so forth, until the card ranked as 20 should be the one which because it is so ambiguous is capable of being seen as expressing many possible emotions.

The instructions for ranking according to the pleasant-unpleasant dimension were:

> I want you to rank each card according to how unpleasant or unhappy the emotion expressed seems to be. That card which is the most unpleasant or unhappy should be ranked 1, the card next most unpleasant should be ranked 2, etc., until 20 should represent the card which is the *least* unhappy or unpleasant, or more simply, that card which is the most pleasant one.

Half of the *S*s received the clarity-lack-of-clarity task first, while the other half received the pleasant-unpleasant ranking first. Subsequent appointments were made by the examiners for the "second part" of the experiment, which occurred from three days to three months later. The "second part" was a retesting for reliability purposes.

Results

Table 1

Correlations Between Pleasantness, Ambiguity, and Emotional Tone Dimensions for Each *S* and for Pooled *S*s

S	Ambiguity vs. Emotional Tone	Pleasantness vs. Emotional Tone	Ambiguity vs. Pleasantness
1	.58	.60	.69
2	.53	.29	.02
3	.22	.56	−.12
4	.58	.50	.60
5	.75	.60	.59
6	.16	.60	−.12
7	.53	.65	.17
8	.52	.50	.21
9	.83	.73	.75
10	.74	.75	.68
11	.23	.70	.09
12	.60	.44	.64
Avg. *r*	.56*	.59*	.40*

* Significant beyond .01 level.

The rho coefficients between the Eron emotional tone-of-story dimension from his study, and in this study the ambiguous-unambiguous, and pleasantness-unpleasantness dimensions of the stimulus properties of the cards are indicated for each *S* in Table 1. This table also shows the average rho intercorrelation between these dimensions for the pooled subjects. Test-retest reliability for the 10 of the 12 *S*s available for retesting, covering an interval of from three days to three months, was .91 for the pleasantness dimension and .84 for the ambiguity dimension. The reliability was not significantly related to the duration of time between the first and second administration of the cards. Individual variation between the 12 *S*s in the correlation of the ambiguous-unambiguous dimension with emotional-tone ranking is considerable (χ^2; $p < .10$) though not for the pleasant-unpleasant vs. emotional-tone correlation (χ^2; $p < .80$). Both of the dimensions (pleasant-unpleasant, and ambiguous-unambiguous) are significantly related to the pooled *S*s' emotional-tone rankings as taken from Eron's study beyond the .01 point for a one-tailed test of significance. The respective correlations are .59, and .56. Hypotheses 1 and 2, therefore, are confirmed. The average correlation between the pleasantness-unpleasantness and ambiguous-unambiguous dimensions is significant at the .01 point with rho = .40. By use of the formula for correction for attenuation, it may be noted that 45% of the "true" variance of the emotional tone ratings is accounted for by the pleasantness-unpleasantness vs. emotional tone-rating correlation, and 44% of the "true" variance is accounted for by the ambiguous-unambiguous vs. emotional-tone rating correlation. The "true" common variance between pleasantness-unpleasantness and ambiguity-unambiguity is only 21%. Hence, both of these dimensions must be taken into account in predicting the emotional tone ratings to TAT cards. Therefore, Hypothesis 3 is seen to be confirmed.

---

[2] The author expresses his thanks to Deldon McNeely, Ronald Pryer, John Roitzsch, and Richard Rolston for their aid in administering the TAT cards.

Lastly, neither order of presentation of the dimensions nor age was found to be related to the size of the correlation for any dimension.

Discussion

It would be helpful in evaluating the findings of this study to utilize a schema of Helson's theory of adaptation-level (3). The theory briefly stated would hold that perception is a resultant of the pooled interaction of the stimulus, background, and residual factors (i.e., personality). Singer (7) has stated that clinicians have been too prone to overweight the importance of personality in determining perception, to the neglect of stimulus and background factors. The results of this experiment seem to support Singer in confirming the existence of several nonprojective determinants of perception on the TAT. The stimulus-determined dimensions of pleasantness and ambiguity together accounted for 61% of the true common variance with the emotional tone-of-stories dimension. Presumably, within the remaining true variance the mechanism of projection occurs, though this mechanism itself has been found to be exceedingly complex (4, 5, 6).

The findings of this study, therefore, would seem to call for a bridling of the unrestrained assumption that, for example, perception of hostility on one of the TAT cards is necessarily indicative of a "hostile" personality. In a technique such as the TAT where the stimulus properties vary considerably in ambiguity and pleasantness from card to card, the perception of hostility on one card might indicate considerable distortion of the stimulus properties (projection?), while on another card reflect only considerable respect for the objective stimulus.

A clinician reviewing the themes of the stories without noting the stimulus characteristics of the card that occasioned it, might draw, therefore, entirely inaccurate inferences. Consequently, it might be beneficial to construct a scale of cards whose pictures differed in pleasantness, structure, etc., by some measured amount based on the construction of judged group norms. In this way, clinicians might come closer to a realization of the suspected correlation between "deviation of perception" and deviation of personality. In short, "deviation of perception" must take cognizance of measured stimulus and background characteristics of the TAT cards before we attribute the remaining variance to deviations in personality patterns.

Summary

The purpose of this study was to examine some of the nonprojective determinants of perception on the TAT.

Three hypotheses were formulated:

1. The emotional tone of stories told in response to the TAT cards is positively related to the pleasantness-unpleasantness of these cards. Those cards pleasant in appearance yield pleasantly toned stories.
2. The emotional tone of the stories is also a function of the structure of the emotional expression on the cards. Those cards most clearly structured (unambiguous) will yield negatively toned stories, while unstructured (ambiguous) cards yield more pleasant themes.
3. While some overlap exists between the dimensions of pleasantness-unpleasantness and ambiguity-nonambiguity, both dimensions show sufficiently low intercorrelation to warrant consideration as separate factors determining the tone of the stories told.

A rho coefficient of .56 between ambiguity-nonambiguity and emotional tone, and .59 between pleasantness-unpleasantness and emotional tone (both significant at the .01 point) seemed to lend support to Hypotheses 1 and 2.

Hypothesis 3 found confirmation in that while both of the dimensions of pleasantness-unpleasantness and ambiguity-nonambiguity correlated $r' = .40$ with each other, each had considerable valid variance in common with the emotional tone dimension which was not accounted for by this correlation.

Received June 14, 1957.

References

1. Eron, L. D. A normative study of the Thematic Apperception Test. *Psychol. Monogr.*, 1950, 64, No. 9 (Whole No. 315).

2. Eron, L. D. Responses of women to the Thematic Apperception Test. *J. consult. Psychol.*, 1953, **17**, 269–282.
3. Helson, H. Adaptation level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, **55**, 297–313.
4. Murstein, B. I. The projection of hostility on the Rorschach, and as a result of ego-threat. *J. proj. Tech.*, 1956, **2**, 418–428.
5. Murstein, B. I. Some comments on the measurement of projection and empathy. *J. consult. Psychol.*, 1957, **21**, 81–82.
6. Murstein, B. I. Studies in projection: A critique. *J. proj. Tech.*, 1957, **21**, 129–136.
7. Singer, J. L. The non-projective aspects of the Rorschach experiment: V. Discussion of the clinical implications of the non-projective aspects of the Rorschach. *J. soc. Psychol.*, 1956, **44**, 207–214.

Journal of Consulting Psychology
Vol. 22, No. 3, 1958

The Validity of the Objective Rorschach and the MMPI[1]

William D. Winter
San Jose State College

and Ramon A. Salcines
Milwaukee County Guidance Clinic

Because of heavy demands for their diagnostic services, psychologists have attempted to develop valid personality tests which are quickly scored, easily evaluated, and adaptable to testing large groups of subjects (*S*s). A number of such rapid screening devices are now in use. One of the newest to be offered to the busy clinician is the Objective Rorschach, a variation of the multiple choice Rorschach, developed by P. O. O'Reilly (1, 3, 5). In this procedure, the *S* is tested individually and he is shown each Rorschach card in turn. Together with each card, he is presented with a list of 12 responses and is asked to choose the two responses which he feels best describe the card. His choices are scored and an over-all measure of disturbance is derived by adding the scores obtained on each card. O'Reilly's data indicate that this procedure is an effective device for distinguishing among normals, neurotics, and psychotics (5), and that the over-all score is also significantly related to clinical improvement in response to therapy (6). The purpose of the present study is to test the validity of the Objective Rorschach when used with a different sample of *S*s, and to compare its ability to identify mental patients with that of a more conventional test, the MMPI (2).

Method

The *S*s for this study consisted of three groups of white males, matched closely as to age and education. Their age range was 17–45 (median 26.6), and their education ranged from 7–17 years of school completed (median 12.2). The first group consisted of 25 normals who were functioning adequately in the community, and who denied ever having had any serious personality difficulties. The second group consisted of 25 normal *S*s who were hospitalized with an acute physical illness, an injury or infection, which as far as could be determined was not obviously psychosomatic in origin. These *S*s were in a stressful situation and experiencing physical and mental discomfort. The third group consisted of 25 hospitalized schizophrenics, all of whom were testable but, in the judgment of the staff, grossly disturbed in their thinking or behavior.

Each *S* was tested by the same psychologist. The test battery included the Objective Rorschach and the group form of the MMPI. The measures of disturbance used in this study are the total Objective Rorschach score and the total score obtained on the MMPI psychosis scale developed by Peterson (7). In scoring this latter scale, the *S* is given one point for each of the following psychotic signs appearing in his MMPI, and the number of points are cumulated:

---

[1] These data were gathered by the second author as part of a master's thesis undertaken at the University of Detroit. The authors wish to express their gratitude to the staffs of the various hospitals and schools, and to the social organizations who cooperated in this study by generously contributing their time and advice.

1. *T* scores over 70 on four or more of the clinical scales
2. *F* greater than 65 (*T* score)
3. *Sc* greater than *Pt*
4. *Pa* or *Ma* greater than 70 (*T* score)
5. *Pa* or *Sc* or *Ma* greater than *Hs* and *D* and *Hy*
6. *D* greater than *Hs* and *Hy*

Peterson suggests that the presence of three or more of these signs indicates psychosis. The ability of the MMPI psychosis scale and the Objective Rorschach to differentiate among our three groups of *S*s was tested by means of chi square.

In analyzing the results of this study, the major emphasis was placed on the ability of these two personality tests to predict psychosis. Specifically, we were interested in whether or not the published cutoff scores on these tests succeeded in separating our schizophrenic patients from the two nonpsychotic groups. In evaluating the predictive efficiency of these tests, it should be kept in mind that only one-third of our sample of 75 *S*s consists of psychotics. Therefore, if one were to arbitrarily diagnose each *S* as not being psychotic, two-thirds of the diagnoses would be correct. For a test to be of value in this situation, its use would have to allow accurate diagnosis of more than the 50 *S*s who could be correctly diagnosed without knowledge of their test results. In addition to the statistical questions of whether or not the three groups obtained significantly different (median) scores on these two tests and whether or not using these tests leads to more correct diagnoses than can be obtained without the tests, one must take into consideration the practical problem of whether the additional number of correct diagnoses one might obtain justifies the time and trouble it takes to administer a given test.

Results

Objective Rorschach

Table 1 shows the distribution of Objective Rorschach scores obtained by our three groups. In his original article, O'Reilly (5) recommends using a cutoff score of 38 and above as an index of psychosis. Although the use of this cutting point leads to the correct classification of 73% of our *S*s, examination of Table 1 will show that only five of our 75 *S*s obtained scores above 37. Although these five were all in the schizophrenic group, 80% of the schizophrenics fall below this cutting point, which would thus appear to be set too high to be of practical value for our sample. A median test showed that our schizophrenic *S*s scored significantly lower on this test than did O'Reilly's psychotic *S*s ($\chi^2 = 18.0$, 1 *df*, $p < .001$). Since O'Reilly does not describe the composition of his sample, we can offer no hypotheses as to why this large difference occurred. Although a higher percentage of correct diagnoses was made by using the test than could be made without it, this increase is insignificant, and the meager results obtained scarcely justify the effort involved in individually testing 75 *S*s.

Further examination of Table 1 will show that a cutoff score of 36 and above, which also correctly classifies 73% of our *S*s, is also of little practical value, since the addition of five correct diagnoses over the 50 obtainable without using the test is not a significant increase. The use of this cutoff results in the correct diagnosis of only nine psychotics, at the expense of incorrectly labelling four non-

Table 1

Distribution of Objective Rorschach Scores

Score	Normal	Physically ill	Psychotic
49			1
43			1
42			1
41			0
40			0
39			2
38			0
37	0	3	0
36	0	1	4
35	1	3	0
34	3	3	1
33	2	1	5
32	2	0	3
31	1	2	2
30	4	2	3
29	1	4	0
28	5	3	0
27	3	1	0
26	3	1	1
25	0	1	1
N	25	25	25

Table 2

*S*s Falling Above and Below the Objective Rorschach Common Median

Score	Normal	Physically ill	Psychotic
32–49	8	11	18
25–31	17	14	7

$\chi^2 = 8.4$; $df = 2$; $p < .02$.

Table 3

*S*s Obtaining Scores of Three or Above on MMPI Psychosis Scale

Score	Normal	Physically ill	Psychotic
3–6	3	5	19
0–2	22	20	6

$\chi^2 = 26.4$; $df = 2$; $p < .001$.

psychotics as being psychotic. Since the two cutting scores discussed above result in the highest percentage of correctly classified *S*s, and since neither proves to be satisfactory, we must conclude that the O'Reilly Objective Rorschach is valueless for the practical problem of making individual diagnoses of psychosis in our *S*s.

To determine whether the Objective Rorschach differentiated our three groups, an extension of the median test was performed (8). Table 2 shows the number of *S*s in each group falling above and below the common median of the entire sample of 75. In this case, the cutoff score is 32 and above, and chi square for this contingency table is 8.4, which for two degrees of freedom is significant beyond the .02 level. Therefore, we can reject the hypothesis that our three groups come from the same population or from populations with equal medians. The Objective Rorschach does have some measure of validity in that the median scores of the three groups differ from one another, although the relationship between test scores and diagnostic category is not high enough to justify the use of this test in making individual predictions of psychosis in a sample such as ours, in which one-third of the *S*s are psychotic. Whether the forecasting efficiency of this test can be greatly improved by item analysis must await further study.

MMPI Psychosis Scale

Peterson's results (7) suggest that a score of three and above on his scale is a conservative yet effective indicator of psychosis. Table 3 shows the number of *S*s in each group who obtained scores above and below this cutting point. By using this cutoff, 81% of our *S*s are correctly classified as to whether or not they are psychotic. This represents a significant increase (*22%*) of 11 correct diagnoses over the 50 that could be correctly determined without the test.

As with the Objective Rorschach, a median test was performed to see whether the medians of the three groups differ. Table 4 presents the number of *S*s in each group who scored at or above the common median (two and above) of the entire sample. The chi square for this table is 18.2 ($p < .001$). Therefore, Peterson's psychosis scale on the group-administered MMPI is more effective than the individually administered Objective Rorschach in differentiating our three groups and in making individual diagnoses of psychosis. However, since a one-point difference in the cutoff score on the Peterson scale makes a distinct difference in the number of correct diagnoses, one might well question the validity of judgments made on the basis of such a short instrument, especially when a given clinical patient obtains a score at or near the cutting point. It should also be noted that both the Peterson scale and the Objective Rorschach drew the majority of their "false positives" from the group of nonpsychotic, physically ill *S*s.

Table 4

*S*s Falling Above and Below the MMPI Psychosis Scale Common Median

Score	Normal	Physically ill	Psychotic
2–6	6	14	21
0–1	19	11	4

$\chi^2 = 18.2$; $df = 2$; $p < .001$.

Table 5

Correlations Between Objective Rorschach Score and MMPI Scales

MMPI scale	r
K	$-.09$
F	$.37^{**}$
$Hs + .5K$	$.22$
D	$.40^{**}$
Hy	$.25^{*}$
$Pd + .4K$	$.29^{*}$
Mf	$-.04$
Pa	$.30^{**}$
$Pt + 1K$	$.37^{**}$
$Sc + 1K$	$.41^{**}$
$Ma + .2K$	$.33^{**}$
Psychosis scale	$.44^{**}$

* $p < .05$.
** $p < .01$.

In Table 5 are listed the product-moment correlations between the Objective Rorschach total score and each of the MMPI scales for our entire sample of 75. There are several low, but significant correlations present, especially between the Objective Rorschach and those MMPI scales indicative of psychosis. The two tests are to some extent measuring the same things, but for the purposes investigated by this study, the MMPI psychosis scale seems to be a more effective instrument than the Objective Rorschach.

Summary and Conclusions

A sample of 25 normal, 25 physically ill, and 25 schizophrenic white male subjects was tested with the O'Reilly Objective Rorschach and the group form of the MMPI. The basic problem investigated is whether the Objective Rorschach or the MMPI psychosis scale developed by Peterson is more effective in differentiating the psychotic *S*s from the other two groups. Our results indicate that although the three groups obtained significantly different medians on the Objective Rorschach, and although this test showed significant correlations with several of the MMPI scales, nevertheless the Peterson MMPI psychosis scale was much more effective in making individual predictions about whether or not a given *S* was psychotic. The Objective Rorschach was of no practical value for this purpose. For our sample, the MMPI psychosis scale was a more valid measure of psychotic disturbance. The MMPI cutoff score suggested by Peterson's results was applicable to our *S*s, but the Objective Rorschach cutoff score suggested by O'Reilly was too high to be useful, since our psychotic *S*s as a group scored significantly lower on this test than did O'Reilly's psychotic sample. Both tests drew most of their "false positives" from the group of physically ill patients. This suggested that a clinician should evaluate the stresses found in a person's life situation before attempting to make a diagnosis of psychosis on the basis of a score on either test.

Although the Objective Rorschach is an interesting and potentially useful screening device, in its present stage of development it does not appear to offer any advantage over the more commonly used and more easily administered and scored MMPI, at least for the purpose investigated in this present study. Further research is needed to cross-validate our findings and to ascertain why our psychotics differed from O'Reilly's sample. It would also be important to determine if the Objective Rorschach is of greater predictive value than tests like the MMPI in such areas as therapeutic improvement (6).

Received June 24, 1957.

References

1. Harrower-Erickson, Molly R., & Steiner, M. E. *Large scale Rorschach techniques.* Springfield, Ill.: Charles C Thomas, 1945.
2. Hathaway, S., & McKinley, J. C. *Minnesota Multiphasic Personality Inventory: Manual.* New York: Psychological Corp., 1951.
3. Kellman, S. Multiple choice Rorschach. In A. Weider (Ed.), *Contributions towards medical psychology.* New York: Ronald Press, 1953. Pp. 625–636.
4. Meehl, P. Profile analysis of the MMPI in differential diagnosis. *J. appl. Psychol.,* 1946, 30, 517–524.
5. O'Reilly, P. O. The objective Rorschach: A suggested modification of Rorschach technique. *J. clin. Psychol.,* 1956, 12, 27–31.
6. O'Reilly, P. O. Therapy: A proposed objective evaluation. *Dis. nerv. Syst.,* 1955, 16, 1–4.
7. Peterson, D. R. The diagnosis of subclinical schizophrenia. *J. consult. Psychol.,* 1954, 18, 198–200.
8. Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Journal of Consulting Psychology
Vol. 22, No. 3, 1958

Age and Sex Influences on MMPI Profile Peak Distributions in an Abnormal Population

Bernard S. Aaronson
New Castle State Hospital, New Castle, Indiana

Sex and age are factors which may have differential influence in bringing people into membership in a population of abnormals. In the case of sex, differences in morphology and physiology, as well as differences in life experiences, may all contribute to differing personality characteristics which cause males and females to be classified as abnormal. In the case of age, different personality factors may help create difficulty for individuals at various stages of the life process. To some extent, such differences may be a function of the demands made upon the individual by our culture as he moves through life. To some extent, particularly among early entrants into abnormal populations, the personality factors themselves may be so maladaptive as to create difficulty regardless of any demands made.

The present study is an attempt to evaluate some of these factors as they affect the incidence of highest points on Minnesota Multiphasic Personality Inventory profiles. Knowledge of the influence of age and sex on these distributions should also help in understanding the significance of given MMPI profile patterns.

Procedure

All of the cases in *An Atlas for the Clinical Use of the MMPI* (4) were considered, and all cases without a codable high point or with any suspicion of invalidity in the profile were eliminated (profile code followed by *X*). Only the first profile code obtained in each case was considered, regardless of how many profiles or how many admissions were recorded. The selection yielded 871 cases of whom 416 were males and 455 females. The sex distribution was then compared with the original sex distribution of 458 males and 510 females in the Atlas sample by chi square, and a chi square of 0.23 was obtained between them, which suggests that they are not significantly different.

The profile codes were divided according to which scale was highest in the profile, and the proportion of the sample accounted for by each such peak was determined. The cases were further divided by age in decades, and the proportion of cases falling in each decade was determined. Because of the small number of cases 60 years old or older, they were combined with the group which fell between 50 and 60 years of age. The expected incidences of profile peaks at given age levels and for males and females were calculated and compared with the actual incidences by chi square. The interactions between sex, age, and profile peak incidence were determined, and further analyses were made to determine at which ages significant relationships occurred and on which scales they took place. Individual incidences were examined to determine the nature and extent of the relationships with age and sex.

Results

The division of the Atlas sample by profile peaks is summarized in Table 1. *D* seems to be the scale on which peaking occurs most frequently. This observation is in accord with Meehl's (6) that *D* is the most frequent peak among abnormals.

Table 2 shows the division of the sample by age. By far the greatest proportion of the sample come into a population of abnormals during the third and fourth decades, after which there seems to be a consistent decline.

Table 1

Incidence of Individuals with Various Profile Peaks in the Atlas Sample

Profile peak	Frequency	Proportion
Hs	77	.09
D	245	.28
Hy	93	.11
Pd	160	.18
Pa	54	.06
Pt	72	.08
Sc	86	.10
Ma	84	.10

Although it is not shown in the table, only 2% of the total sample were 60 or older, and most of these were very close to 60.

When the interaction of sex and profile peak incidence is tested by chi square, a value of 39.99 is obtained ($p < .001$). Table 3 shows that profiles with *Hs* as the highest point tend to be obtained by men ($p < .05$). Profiles with *Pt* as the highest point also tend to be obtained by men ($p < .02$). Profiles with *Hy* as the highest point tend to be obtained by women ($p < .001$). A trend also seems present for women to obtained profiles in which *Pa* is highest, but it does not attain significance.

Comparison of the obtained sex distributions for each profile peak shown in Table 3 with one another by chi square yields much the same results. Significantly more men seem to obtain profiles with *Hs* as the highest point than obtain profiles with *D, Hy,* or *Pa.* These differences are significant at beyond the .05, .001, and .01 levels, respectively. More men obtain profiles with *Pt* as the highest point than obtain profiles with *Hy* or *Pa.* These differences are significant at the .001 and .01 levels, respectively. Significantly more women obtained profiles with *Hy* as the highest point ($p < .001$) than obtain profiles with any other scale except *Pa* as their peak. More women obtain profiles with peaks on *Pa* than obtain profiles with peaks on *Hs, Pt, Sc,* and *Ma.* The first three comparisons are significant at the .01 level, the comparison with *Ma* is significant at the .05 level.

Table 2

Incidence of Individuals of Various Age Levels in the Atlas Sample

Age level	Frequency	Proportion
10–19	135	.16
20–29	253	.29
30–39	244	.28
40–49	140	.16
50–75	99	.11

When the interaction between age and incidence of profile peaks is tested, a chi square of 102.71 is obtained, which is significant at the .001 level. The data in Table 4 suggest that significant age relationships occur with *Hs, D, Pd,* and *Sc.* The relationship between *Hs* and age is significant at the .01 level; the latter three are significant at the .001 level. Significant relationships between profile peaks and age are noted at every period studied, except that between 30 and 39.

Inspection of the expected and obtained incidences of profile peaks at the various age levels as shown in Table 4 suggests that *Hs* and *D* peaks are comparatively uncommon in early life but become much more common as the individuals entering the clinical population become older. The reverse seems to be true of *Pd* and *Sc.* In addition, *Pt* peaks seem to be more prominent in the group which enters into an abnormal population in the twenties.

Table 3

A Comparison by Chi Square of Expected and Obtained Distributions of Men and Women Obtaining Given Highest Points on the MMPI Profile

Scale	Expected *N*		Obtained *N*		Chi square
	M	*F*	*M*	*F*	
Hs	37	40	47	30	5.20[a]
D	115	130	114	131	0.02
Hy	44	49	21	72	22.82[c]
Pd	75	85	82	78	1.23
Pa	25	29	19	35	2.68
Pt	34	38	44	28	5.57[b]
Sc	40	46	44	42	0.75
Ma	39	45	45	39	1.72

[a] p is at least .05.
[b] p is at least .02.
[c] p is at least .001.

Table 4

Expected and Obtained Incidences of MMPI Profile Peaks in Successive Decades

Peak scale	Age range									
	10–19		20–29		30–39		40–49		50–59	
	E	O	E	O	E	O	E	O	E	O
Hs	12	5[a]	23	13[a]	22	29	12	19[a]	9	11
D	39	15[d]	71	60	68	75	39	51[a]	27	44[d]
Hy	16	12	28	23	27	26	16	22	10	10
Pd	25	46[d]	45	57	44	37	25	11[c]	17	9[c]
Pa	8	7	15	14	15	22	8	6	6	5
Pt	11	7	20	31[b]	19	18	11	13	8	5
Sc	14	27[c]	25	28	24	17	14	8	10	6
Ma	14	16	25	30	24	20	14	14	10	10

[a] Difference is significant at the .05 level.
[b] Difference is significant at the .02 level.
[c] Difference is significant at the .01 level.
[d] Difference is significant at the .001 level.

Discussion

The close association between *Hy* as the highest point and being female suggests the early theories of hysteria as a feminine complaint (2, p. 370). The association of peaks on *Sc* with youth suggests the Kraepelinian formulation of dementia praecox as a disorder of the young (9, p. 21). The latter is also in accord with the association of the diagnosis of schizophrenia with youth among patients admitted to mental hospitals (7).

When the description of the scales as given in the manual (3) for this test is taken into account, men seem to predominate among those with vague physical complaints which do not seem to be obviously successful in getting them out of unacceptable situations. They also seem to include a greater number of individuals with obsessive-compulsive complaints, phobias, and excessive worry and doubt. Women seem to predominate among individuals with specific somatic complaints which may be alleviated easily but return under stress. They seem to show more pathology in the area of interpersonal relationships. It should be noted that the scales on which men peak are among the scales which Welsh (8) uses to suggest internalization, while *Hy*, on which women particularly tend to predominate, is among the scales used by him to suggest externalization. The data suggest that men may be given to internalization of their problems while women may tend to externalize them.

With regard to age, the scale description in the manual suggests that younger individuals coming into an abnormal population may be characterized by difficulty in their relations with the environment as typified by acting out against it or withdrawing from it. Older persons seem to have more problems with their own physical and mental functioning. In this connection, it is of interest to note that *Hs* and *D*, which are common in the older age group, are among the scales regarded by Welsh (8) as suggestive of internalization of problems, while *Pd*, which occurs among the young, is one of the scales used by him to suggest externalization. The significantly greater proportion of individuals with peaks on *Pt* is in accord with a previous finding by Brozek (1) that younger males tend to admit to more obsessive-compulsive symptoms than older males. Brozek, however, did not find significant differences on the *Hs* scale. It should be noted that Brozek was comparing absolute scale magnitudes, while this study deals with the relative position of the scale in the profile. Lauber and Dahlstrom (5) have pointed out that the utilization of absolute scale scores may obscure relationships which the utilization of profile codes may reveal.

When the data on age and sex are considered together, *Hs* and *Pt* which occur among males also seem to show a relation with age. Since the scales on which women peak show no such relationship, this might suggest that men tend to enter into clinical populations in the twenties and forties, while women show no tendency to enter at any particular age.

Summary

An analysis of the distribution of the profile peaks of 871 cases taken from *An Atlas for the Clinical Use of the MMPI* with regard to the sex and age distributions of the sample suggests that men tend to obtain peaks on *Hs* and *Pt* and women on *Hy* and *Pa*. Peaks on *Pd* and *Sc* seem characteristic of a younger age group and peaks on *Hs* and *D* of an older. Some possible implications of this are noted. The scales on which women peak tend to imply a more social orientation than those on which men peak. Younger entrants into an abnormal population seem characterized by behavior patterns maladaptive to societal demands. Older entrants seem characterized by problems centering around their own physical and mental functioning.

Received June 26, 1957.

References

1. Brozek, J. Personality changes with age: An item analysis of the Minnesota Multiphasic Personality Inventory. *J. Gerontol.*, 1955, **10**, 194–206.
2. Dorcus, R. M., & Shaffer, G. W. *Textbook of abnormal psychology.* (3rd ed.) Baltimore: Williams & Wilkins, 1945.
3. Hathaway, S. R., & McKinley, J. C. *Minnesota Multiphasic Personality Inventory manual* (rev. ed.). New York: Psychological Corp., 1951.
4. Hathaway, S. R., & Meehl, P. E. *An atlas for the clinical use of the MMPI.* Minneapolis: Univer. of Minnesota Press, 1951.
5. Lauber, Margaret, & Dahlstrom, W. G. MMPI findings in the rehabilitation of delinquent girls. In S. R. Hathaway & E. D. Monachesi (Eds.), *Analyzing and predicting juvenile delinquency with the MMPI.* Minneapolis: Univer. of Minnesota Press, 1953.
6. Meehl, P. E. Profile analysis of the MMPI in differential diagnosis. *J. appl. Psychol.*, 1946, **30**, 517–524.
7. National Committee Against Mental Illness, Inc. *What are the facts about mental illness?* New York: Author, 1957.
8. Welsh, G. S. An anxiety index and an internalization ratio for the MMPI. *J. consult. Psychol.*, 1952, **16**, 65–72.
9. White, R. W. *The abnormal personality.* New York: Ronald Press, 1948.

Journal of Consulting Psychology
Vol. 22, No. 3, 1958

Spearman Correlations Between MMPI Scores of College Students and Their Parents[1]

Andrew L. Sopchak[2]

Memorial Center for Cancer and Allied Diseases, New York, New York

The purpose of this study was to determine the relation between MMPI scores for college students and their parents. As part of their laboratory work in elementary psychology, students at the University of Maine were given the booklet form of the MMPI. The parent(s) of 25 male students and 25 female students volunteered to take the MMPI.

The tests were scored according to standard procedure and tests whose validity scores on *?*, *L*, and *F* were high, were discarded. The rank correlations are based on raw scores on the MMPI for students and their parents.

The mean age for the 50 students was 19.9 years. The 41 mothers had a mean age of 47.7 years, and the 35 fathers had a mean age of 50.5 years.

The mothers' education ranged from high school graduates to normal school training or college graduates. Some of the fathers had only an eighth-grade education, while others were college graduates, holders of master's degrees, or were dental and medical school graduates. The mothers were mostly housewives. The father were painters, lobster men, managers, engineers, dentists, and physicians.

Results

Tables 1 and 2 present the ranges, means, and standard deviations on the MMPI scales for the students and their parents. The scores of the students are similar to scores presented by other investigators (1, 2, 5, 6, 7, 8) on other subjects. Although the present samples are relatively small, the mean scores are about the same as those obtained by Sopchak (7) on a much larger sample from the same geographical area.

In comparison to Schmidt's (5) adult men, both of the groups of fathers are considerably different. However, Schmidt's subjects were mental hospital patients. They are also different from the original standardization group of McKinley and Hathaway (3). In comparison to Verniaud's (8) adult women, both groups of women are similar to hers except on the "neurotic triad" scales on which our women are considerably higher. The means for the fathers and mothers are very similar. The mean scores of the fathers and mothers are more nearly alike than are the scores of the subjects with their parents.

Table 3 presents the Spearman correlations between the scores for the subjects and the scores for their parents.

More significant correlations were obtained between the scores of males and fathers than between those of males and mothers. On the other hand, many more significant correlations between the scores of females and their mothers than between females and their fathers were obtained. It appears that males "identify" with their fathers more than with their mothers, while females "identify" much more with their mothers than with their fathers.

Five significant correlations were obtained between the males and their fathers, but only three between the females and their fathers. Only three significant correlations were obtained between the males and their mothers, but nine between the females and their mothers.

---

[1] This study was supported in part by Research Grant M-884(C) from the National Institute of Mental Health, National Institutes of Health, US Public Health Service.

[2] Formerly at the University of Maine where these data were obtained.

Table 1

Ranges, Means, and Standard Deviations in *t*-Score Units for 25 Males and Their Parents

Group		*Hs*[a]	*D*	*Hy*	*Pd*[a]	*Mf*	*Pa*	*Pt*[a]	*Sc*[a]	*Ma*[a]
Men (25)	Range	42–88	33–80	43–80	29–79	12–82	33–67	9–93	44–101	38–73
	Mean	52.7	53.4	56.0	53.9	60.0	51.5	56.7	57.6	58.4
	SD	8.5	10.7	8.9	10.4	14.6	9.1	15.8	12.0	8.8
	% above 70	4	8	4	8	24	0	8	16	16
Fathers (20)	Range	36–86	34–89	44–80	29–69	41–65	41–65	36–72	32–73	35–65
	Mean	56.9	55.6	56.5	52.0	54.1	52.1	53.0	50.2	50.5
	SD	12.6	11.4	9.9	9.6	8.9	8.9	10.0	10.9	9.9
	% above 70	15	5	10	0	0	0	5	5	5
Mothers (21)	Range	38–82	44–78	42–72	32–76	28–95	35–62	23–73	38–69	27–65
	Mean	56.3	59.5	57.0	52.3	51.8	51.0	51.5	50.4	44.9
	SD	9.8	8.1	8.7	11.5	19.4	7.5	12.9	9.0	10.5
	% above 70	9	9	9	9	19	0	9	0	0

[a] Corrected with the *K* factor.

Table 2

Ranges, Means, and Standard Deviations in *t*-Score Units for 25 Females and Their Parents

Group		*Hs*[a]	*D*	*Hy*	*Pd*[a]	*Mf*	*Pa*	*Pt*[a]	*Sc*[a]	*Ma*[a]
Women (25)	Range	42–63	32–69	42–64	30–79	32–82	35–65	22–69	40–95	30–78
	Mean	49.6	47.8	55.6	54.0	52.0	54.1	51.7	57.1	56.0
	SD	6.3	9.0	5.3	10.0	10.9	7.4	10.9	12.0	10.4
	% above 70	0	0	0	8	4	0	0	12	12
Fathers (18)	Range	43–60	36–70	45–71	39–81	45–78	42–70	30–79	41–65	26–78
	Mean	52.3	53.8	56.5	54.7	57.0	53.5	52.6	54.3	51.0
	SD	6.6	10.3	7.3	11.7	8.7	8.3	9.8	9.1	14.4
	% above 70	6	6	6	17	17	11	11	6	11
Mothers (25)	Range	39–77	36–77	38–73	36–69	28–87	38–68	32–73	41–76	33–65
	Mean	51.5	55.0	57.0	53.6	48.7	52.8	51.0	51.9	50.0
	SD	7.5	11.0	7.0	7.2	12.4	7.4	12.2	8.7	8.0
	% above 70	4	16	4	0	8	0	4	4	0

[a] Corrected with the *K* factor.

Table 3

Spearman Correlations Between Scores for Students and Scores for Their Parents on the MMPI

Groups compared	*N*	*Hs*	*D*	*Hy*	*Pd*	*Mf*	*Pa*	*Sc*	*Pt*	*Ma*
Males and fathers	20	.19	−.09	−.38*	.41*	.18	.42*	.37*	.47*	.04
Males and mothers	21	.05	−.20	.06	.18	−.15	.46*	.48*	.60**	.22
Females and fathers	18	.38	.47*	.20	.30	.13	.47*	.26	.65**	.24
Females and mothers	25	.75**	.64**	.65**	.67**	.56**	.61**	.78**	.73**	.50**
Fathers and mothers of males	16	.58*	.17	.54*	.33	−.25	.49*	.53*	.36	.55*
Fathers and mothers of females	18	−.09	.01	−.20	.28	.38	.06	−.08	.14	.28

* Significant at .05 level.
** Significant at .01 level.

There are more significant correlations between the subjects and both parents on the "psychotic triad" scales than on the "neurotic triad" scales. However, scores on all three "neurotic triad" scales are significant for the females and their mothers. A most impressive finding is the obtaining of correlations significant at the .01 level on each scale of the MMPI between the females and their mothers. An explanation for this phenomenon is not offered presently.

Five significant correlations were obtained between the fathers and mothers of males, but none between the fathers and mothers of females. The significant correlations were on the "neurotic" and "psychotic triad" scales. Why the parents of males are different from the parents of females is difficult to explain. This observation should provide a lead for future studies.

According to Sakoda *et al.* (4), out of 54 calculated statistics it is possible to obtain 6 significant at the .05 level and 3 significant at the .01 level by chance. Fourteen correlations significant at the .05 level and 11 significant at the .01 level were obtained. Almost three times as many significant correlations were obtained as could have resulted by chance. It is, therefore, very unlikely that the significant correlations obtained are due to chance factors alone.

Summary and Conclusions

Over a period of four years, the group form of the MMPI had been administered to many students in elementary psychology. The parents of 25 men and 25 women volunteered to take the MMPI. This study reports on the relationship between the MMPI scores for these students and their parents.

The results permit the following conclusions:

1. The subjects and their parents are "normal" according to the criteria of the authors of the test.

2. The scores of the fathers and mothers are different from the scores of men and women in the original standardization group.

3. The scores for fathers and mothers are more similar to each other than they are to the scores of the subjects.

4. It appears that the fathers of women are more variable than the fathers of men. The mothers of men are more variable than the mothers of women. This is an interesting and challenging observation which deserves further investigation.

5. There are more significant correlations between scores of males and fathers and scores of females and mothers than between scores of males and mothers and females and fathers.

6. Significant correlations on the "psychotic triad" scales were obtained between the male and female subjects and both parents.

7. On the "neurotic triad" scales, the largest number of significant correlations was between females and their mothers.

8. Correlations significant at the .01 level were obtained for each of the nine MMPI scales between the scores of females and their mothers. No explanation for this unusual finding was offered.

9. Scores for fathers and mothers of males correlated significantly on five MMPI scales.

10. Scores for fathers and mothers of females did not correlate significantly on any of the scales. The reason for the discrepancy in correlations between parents of males and parents of females remains unexplained.

Received June 17, 1957.

References

1. Brown, H. Similarities and differences in college populations on the Minnesota Multiphasic Personality Inventory. *J. appl. Psychol.*, 1948, **32**, 541–549.
2. Lough, Orpha M. Teachers college students and the Minnesota Multiphasic Personality Inventory. *J. appl. Psychol.*, 1946, **30**, 241–247.
3. McKinley, J. C., & Hathaway, S. R. *Manual for the Minnesota Multiphasic Personality Inventory.* New York: Psychological Corp., 1951.
4. Sakoda, J. M., Cohen, B. H., & Beall, G. Test of significance for a series of statistical tests. *Psychol. Bull.*, 1954, **51**, 172–175.
5. Schmidt, H. O. Test profiles as a diagnostic aid: The Minnesota Multiphasic Inventory. *J. appl. Psychol.*, 1945, **29**, 115–131.
6. Sopchak, A. L. Parental "identification" and "tendency toward disorders" as measured by the Minnesota Multiphasic Personality Inventory. *J. abnorm. soc. Psychol.*, 1952, **47**, 159–165.
7. Sopchak, A. L. College student norms for the Minnesota Multiphasic Personality Inventory. *J. consult. Psychol.*, 1952, **16**, 445–448.
8. Verniaud, Willie M. Occupational differences in the Minnesota Multiphasic Personality Inventory. *J. appl. Psychol.*, 1946, **30**, 604–613.

Journal of Consulting Psychology
Vol. 22, No. 3, 1958

The MMPI in Cerebral Palsy

Tom Linde and C. H. Patterson[1]
University of Illinois

The problem of whether there are personality characteristics associated with the various disabilities, physical and psychosomatic, has been the object of considerable attention (1). Among others, there has been discussion of the ulcer personality, the tuberculosis personality, and the multiple sclerotic personality. The present study approaches the study of the personality of individuals with cerebral palsy, through use of the Minnesota Multiphasic Personality Inventory. It is the opinion of the authors that personality types associated with specific disabilities do not exist. Harrower and Herrmann (5), reviewing studies of multiple sclerotic patients concluded that there is no evidence of a multiple sclerosis personality. There have been no adequate objective studies of those with cerebral palsy.[2]

Sample

It is estimated that there are 550,000 cerebral palsied people in the United States, about half of them unable to work because of the severity of the disability. It was felt that if a cerebral palsy personality existed, it would manifest itself in those most severely disabled. The sampling of handicapped or disabled persons, even though they are numerous, is exceedingly difficult. Usually, intact groups or selected individuals are most conveniently available. The present sample is not a random sample, but it is felt to be rather representative of adult, severely disabled, cerebral palsied individuals, without mental or intellectual impairment, in the midwest.

The sample includes 14 males and 19 females, ranging in age from 20 to 43 years, with a mean and median age of 24. Twenty had a speech defect; nine were confined to wheel chairs all the time and three part of the time; three used crutches, and the remaining seventeen were ambulatory. Only three had full use of both hands, the remainder having at least some use, however. The range in education was from four to sixteen years, the mean being 11.4 and median 12 years. Twelve had been educated in the public schools entirely, the remainder in orthopedic or a combination of public and orthopedic schools except for three who had home instruction.

The socioeconomic background of the subjects, as indicated by the occupations of the parents, was wide, ranging from unskilled labor to engineer, though apparently concentrating in the middle occupational levels.

Results

Males and females were not significantly different, except for the *Pt* scale where the difference was significant at the .05 level.[3] The sexes were pooled for further comparisons with other groups containing both sexes.

Table 1 contains the means and standard deviations for the present sample, together with the results of statistical comparisons with (*a*) the MMPI norm group, (*b*) Canter's (3) multiple sclerosis group, and (*c*) Shontz' (7) multiple sclerosis group. Com-

---

[1] From the MA thesis of the senior author, done under the direction of the junior author.

[2] For a review of studies of the personality of the cerebral palsied, see the senior author's MA thesis, entitled "Personality elements of thirty-three adults with cerebral palsy as measured on the Minnesota Multiphasic Personality Inventory," on file in the University of Illinois library.

[3] All comparisons were preceded by an F test for homeogeneity of variances. Where variances were significantly different, the Cochran-Cox test of the significance of differences between means was applied. An asterisk indicates instances in which this test, rather than the usual t test, was applied.

Table 1

Means and Standard Deviations on MMPI Scales for 33 Cerebral Palsied Adults, and Statistical Results of Comparisons with Other Groups

Scale	33 CPs Mean	Comparison with manual norms $N = 800$		Comparison with Canter's MS $N = 33$		Comparison with Schontz' MS $N = 16$	
		t	*p*	*t*	*p*	*t*	*p*
?	32.00						
L	51.00						
F	49.00						
K	51.80						
Hs	55.55	4.14*	.01	9.90*	.01	12.65*	.01
D	52.27	1.45	>.05	10.05*	.01	4.24*	.01
Hy	56.48	4.44	.01	8.40	.01	3.53	.01
Pd	54.36	2.45	.02	1.80	>.05	.52	>.05
Mf	54.67	2.39	.02	.33	>.05	.21	>.05
Pa	55.27	3.23	.01	.91	>.05	.83	>.05
Pt	56.36	3.48	.01	2.01*	>.05	.91	>.05
Sc	63.36	8.25	.01	.08*	>.05	1.35	>.05
Ma	60.76	4.11	.01	2.19	>.05	1.55	>.05

* Cochran-Cox test used. Others are *t* tests.

parisons with the latter two groups were made by reconstructing variances from *SD*s given in the publication by Canter and the extended report by Shontz.

Discussion

The comparisons with the standardization sample indicate that the sample of CPs studied was significantly higher on all but the *D* scale. The actual differences are relatively small, however, except on the *Sc* and *Ma* scales, where they reach 10 or more *T*-score points. Before ascribing importance to these results, however, two considerations are suggested. First, the scores of the CPs do not differ much, if any, from scores reported on college students (2, 4, 8), except on the *Sc* scale. It might be an interesting hypothesis worthy of exploration that college students and the handicapped, perhaps for differing reasons (in the one case superior intelligence and in the other the disability) have something in common contributing to some general elevation on the MMPI scales, whether this be a higher sensitivity or insight, or some other factor.

Secondly, an analysis of the answers to the eighteen items contributing solely to the *Sc* scale was made for 24 subjects (*S*s) for whom the MMPI answer sheets were available. Seventeen answered the item: "I have never been paralyzed or had any unusual weakness of any of the muscles" in an abnormal manner. Nine responded to the item: "Sometimes my voice leaves me or changes even though I have no cold" abnormally; eight responded to the item: "I refuse to play some games because I am not good at them" abnormally. This suggests that these, and other handicapped or disabled *S*s, raise their scores on the MMPI because of responses that are classified as abnormal but are in fact true for them. On the other hand, the disabled may, again because of factors related to their disabilities possibly, manifest evidences of actual social maladjustment or withdrawal. Ten *S*s responded to the item: "The things that some of my family have done have frightened me" abnormally; and seven responded to the item: "I have never been in love with anyone" in the deviant manner. On the other hand, only three felt that they did not make friends about as quickly as others do.

The comparisons with the multiple sclerotic samples suggest that the CPs are better adjusted, or less neurotic, than the MSs. It might well be expected that the CP, whose disability is congenital, may be better adjusted to his disability than the multiple sclerotic, whose disability has had a relatively recent onset, and is unpredictable in progression, leading to the possible development of neurotic concern and anxiety (5, p. 10). Shontz (7) found that his MS *S*s were significantly higher on *Hs* and *Hy* than a chronic illness group.

While there are thus statistically significant differences between the CPs in this study and the MMPI norm group, and in other studies between *S*s with other physical handicaps or disabilities and the test norms, the conclusion that there are specific personality types or patterns associated with particular disabilities does not appear to be warranted. Scores on tests which, like the MMPI and other questionnaire tests of personality, include many items related to physical or psychosomatic conditions are probably elevated by reason of the actual presence of physical defects or complaints. Furthermore, it appears that there is, as yet at least, no evidence of unique personality patterns or types associated with differing disabilities. It is suggested that, apart from elevations due to the actual presence of physical defects or disabilities, there may be a common tendency toward personal-social maladjustment among all handicapped individuals. It is generally agreed that a disability can lead to adjustment problems. The disability may be a source of anxiety and/or a tendency toward withdrawal from social contacts. To some extent, both of these are normal in the handicapped. Whether the deviation is greater than can be expected is a question which probably should not be answered by comparing the handicapped individual with the norms for a nonhandicapped group.

Received July 12, 1957.

References

1. Barker, R. G., Wright, Beatrice A., Meyerson, L., & Gonick, Mollie R. *Adjustment to physical handicaps and illness: A survey of the social psychology of physique and disability.* Bulletin No. 55 (rev.) New York: Social Science Research Council, 1953.
2. Brown, H. L. Similarities and differences in college populations on the MMPI. *J. appl. Psychol.,* 1948, **32**, 541–549.
3. Canter, A. H. MMPI profile in multiple sclerosis. *J. consult. Psychol.,* 1952, **15**, 353–356.
4. Goodstein, L. D. Regional responses among male college students on the MMPI. *J. consult. Psychol.,* 1951, **18**, 437–445.
5. Harrower, Molly R., & Herrmann, Rosalind. *Psychological factors in the case of patients with multiple scleroses: for use of physicians.* New York: National Multiple Sclerosis Society, 1953.
6. Hathaway, S. R., & McKinley, J. C. *Manual for the Minnesota Multiphasic Personality Inventory.* New York: Psychological Corp., 1951.
7. Shontz, F. C. MMPI responses of patients with multiple sclerosis. *J. consult. Psychol.,* 1955, **19**, 74.
8. Sopchak, A. L. College student norms for the MMPI. *J. consult. Psychol.,* 1952, **15**, 445–448.

Journal of Consulting Psychology
Vol. 22, No. 3, 1958

An Evaluation of the Comparability and Reliabilities of Two Behavior Rating Scales for Mental Patients[1]

Donald W. Stilson,[2] Donald J. Mason, Malcolm D. Gynther, and Boris Gertz

South Carolina State Hospital and the University of South Carolina

A problem frequently encountered in mental health research is that of accurately and economically evaluating the level of adjustment of hospitalized mental patients (1). Two commonly used criteria of adjustment are discharge rates and transfer to "improved" wards. These criteria suffer from several inadequacies which limit their usefulness. For example, these measures (*a*) usually are dependent upon clinical (subjective) impressions, (*b*) are often influenced by interest and/or pressure from relatives, (*c*) depend upon available facilities, (*d*) are frequently conceived as being therapeutic interventions themselves, (*f*) are seldom repeatable on the same patients, (*g*) are not applicable to certain classes of mental patients (e.g., security problems, the aged, the physically infirm), (*h*) are not consistently used from one hospital to another, and (*i*) may be related to changes in administrative policies.

Two instruments which have been designed to provide a convenient method for evaluating patient adjustment are the Multidimensional Rating Scale for Psychiatric Patients (2) and the Hospital Adjustment Scale (3). Hereafter, these scales will be designated as the MRSPP and the HAS, respectively. These instruments are scored on the basis of comparatively brief patient observation periods and are relatively free of the inadequacies described above for criteria such as discharges and transfers. The present study is aimed at determining the degree to which ratings on these two scales are comparable, as well as answering several questions regarding the test-retest reliability of the MRSPP. The administration of the MRSPP requires highly trained personnel, while the HAS can be used by ward personnel with relatively limited training. Therefore, in most instances, it would be more economical to use the HAS if the measurement of patient adjustment obtained on the two instruments are highly correlated.

The degree to which the two instruments correlate with one another is, of course, a function of the reliabilities of each of the instruments. A further purpose of the present study is to assess the reliability of these scales. Many of the practical applications of the MRSPP would require repeated measurements (of the same patient group) separated by some period of time. Such measurement would presumably be sensitive to whatever changes in patient behavior had been produced by treatment. Yet, differences in ratings might be a product of a number of factors quite unrelated to the effects of therapy. For example, the test-retest reliability of the instrument may vary as a function of the type and degree of pathology shown by the patient. A cyclic manic-depressive may exhibit far greater day-to-day variation in behavior than would a severely regressed chronic schizophrenic. Similarly, the amount of personality disorganization demonstrated by any given schizophrenic might influence the reli-

---

[1] The authors wish to express their gratitude to Martha Bradley, Director of Nursing Education at South Carolina State Hospital, for her numerous contributions to the execution of this research. Thanks are also extended to Ruth A. Gynther, Elmore A. Martin, Jr., and Charles Presher who aided in rating patients and assembling data.

[2] Now at the University of Colorado.

ability with which he could be rated. Another factor in evaluating the test-retest reliability of an instrument is whether or not the same rater makes both ratings. Aside from the obvious chance of contaminated ratings due to a rater's recall of his previous scorings, better rapport might be established as a function of the patient and rater having discussed the same materials earlier.

To summarize, the present study attempts to assess the correlation between HAS scores and morbidity scores of the MRSPP and to evaluate the test-retest reliability of the MRSPP as a function of: (*a*) whether same or different raters make the pre- and post-ratings and (*b*) the degree of morbidity exhibited by the patients as measured by HAS scores.

Procedure

Subjects

Thirty-six *S*s were drawn from a group of 240 chronic psychotic, female patients 23 to 55 years of age. All had been diagnosed "schizophrenic" by the psychiatric staff. The group from which the experimental sample was drawn was housed in two buildings, each of which consisted of four wards. Patients were assigned to wards according to some relatively crude evaluation of morbidity. Patients on Ward A, for example, were generally less aggressive, more communicative, and showed fewer signs of severe regression than those on Ward D. Thus, a stratified population was defined with HAS total scores, building, and wards forming the three strata.

Use of the HAS

The HAS total scores obtained by student nurses approximately one month before the beginning of the present study were available for each of these patients. Before their attempt to rate members of the experimental sample, the 57 student nurses were given three two-hour training sessions and two practice ratings with patients not in the experimental sample. In the initial training, the meaning of each item on the HAS was discussed and clarified. The student nurses were then divided into groups of from three to ten members. Each of the members of a group rated the same patient whom she had observed for at least one month. In the training session following each practice rating, items showing less than 75% agreement between raters were discussed, and rules regarding the use of these items were prescribed. The distribution of the HAS total scores on the experimental sample was then divided into equal thirds. The first group consisted of scores of 33 or less; the second, of scores from 34 to 74, and the third of scores of 75 and over.

Use of the MRSPP [3]

Seven psychologists served as raters on the MRSPP. Two had the M.A. degree with three and seven years of clinical experience. The remaining five raters had doctoral degrees and have had from one to twelve years of clinical experience. During a training period, each rater administered the scale at least three times, with two or more raters evaluating the same patient. Following each rating, the results for the different raters were compared item by item, and those items which most frequently yielded disagreement were discussed by all seven raters. Specific "conventions" were adopted concerning those items which most frequently showed disagreement in an attempt to reduce interrater variability.

The interview procedure was standardized with respect to (*a*) the opening comments of the rater, (*b*) the use of certain questions which it was believed might elicit responses which were relevant to items on the MRSPP, and (*c*) questions regarding specific information (e.g., orientation) which was necessary to complete the ratings.

Patients were interviewed about one-half hour for each rating. The interviews were conducted in private rooms on the ward. An aide or nurse was then interviewed in order to obtain the information required to complete the last 22 items of the MRSPP. The same aide or nurse was always interviewed in obtaining test and retest information for a given patient.

---

[3] The rating conventions and interviewing guide have been deposited with the American Documentation Institute. Order Document No. 5538, remitting $1.25 for microfilm or $1.25 for photocopies.

Experimental Design

The basic design of the experiment was a 2 × 2 × 3 factorial in which the variables were: (*a*) same or different rater on test and retest, (*b*) test versus retest scores, and (*c*) low, medium, and high HAS scores. Six patients were assigned to each of the 12 cells on the basis of the population stratification mentioned earlier. Each patient appeared in both test and retest categories. The samples were drawn randomly within each combination of the population strata, with sample probability proportional to the population frequency.

No rater was paired with another rater more than once under the "different rater" condition, and each rater obtained approximately half his ratings under the "test" condition. Insofar as possible, each rater obtained an equal number of ratings from each ward and each building. "Test" and "retest" ratings were obtained three weeks apart.

Results and Discussion

The data obtained in this study are summarized in Table 1. Inspection of this table reveals that the only significant main effect was HAS score ($p < .01$). This indicates a statistically significant relationship between the MRSPP morbidity scores and HAS total scores. The Pearson product-moment correlation between these two variables is − .57. The HAS-MRSPP correlation corrected for attenuation was − .72. The negative value refers to a positive relationship as high HAS scores, and low MRSPP scores are interpreted as signifying good adjustment. When test and retest scores were obtained two weeks apart using student nurse raters, the reliability of the HAS was estimated to be .79. With a three-week period intervening, the MRSPP morbidity scores yielded an .80 reliability coefficient. All three of these correlations are significant beyond the .01 level of confidence. The fact that the Rater and Test-Retest variables were not significant suggests that: (*a*) it makes no difference whether same or different raters obtain pre- and postmeasures, and (*b*) the MRSPP is sufficiently reliable to warrant its use in evaluation studies requiring repeated measurements on the same patients.

Table 1

Summary of the Analysis of Variance

Source	*df*	m.s.	*F*
HAS	2	3708.88	13.90
Test-Retest	1	4.50	0.02
Rater	1	58.67	0.22
HAS by Test-Retest	2	19.73	0.07
HAS by Rater	2	304.55	1.14
Test-Retest by Rater	1	14.22	0.05
Test-Retest by HAS by Rater	2	43.87	0.16
Within cells	60	266.82	
Total	71	341.42	

Further inspection of Table 1 indicates that none of the interactions approach significance at the .05 level. The fact that the HAS by Test-Retest interaction is not significant suggests that reliability of group means is not a function of level of morbidity as measured by the HAS. The lack of significance for the HAS by Rater interaction casts doubt on the hypothesis that repeated interviewing of the same patient has an effect on the degree of morbidity inferred by the examiner. The finding of nonsignificance for the Test-Retest by Rater interaction demonstrates that reliability is not contingent on whether the same rater obtains both test and retest ratings.

Summary

This experiment was devised to evaluate the utility of two scales widely used to rate patient morbidity. A further consideration was the reliability of each of these scales. In addition, the test-retest reliability of the MRSPP was studied as a function of degree of patient morbidity and of same or different rater on test and retest.

Estimates of test-retest reliability of the HAS and MRSPP morbidity scores were .79 and .80, respectively. The correlation between these two scales was found to be − .57. Further findings with the MRSPP were: (*a*) same or different raters may be used for obtaining "pre-" and "post" measures, (*b*) reliability of group means is unrelated to level of morbidity, and (*c*) rating a patient more than once has no effect on the degree of morbidity inferred.

Both the MRSPP and the HAS were found to be reliable instruments for repeated ratings of psychiatric patients. For those interested in economy, the feasibility of using nonprofessional raters with the HAS is of considerable interest.

Received June 20, 1957.

References

1. Lorr, M. Rating scales and check-lists for the evaluation of psychopathology. *Psychol. Bull.*, 1954, **51**, 119–127.
2. Lorr, M., Jenkins, R. L., & Holsopple, J. Q. Multidimensional scale for rating psychiatric patients, hospital form. *U. S. Vet. Adm. tech. Bull.*, TB 10–507, Nov. 16, 1953.
3. McReynolds, P., Ballachey, E. L., & Ferguson, J. T. Development and evaluation of a behavior scale for appraising the adjustment of hospitalized patients. *Amer. Psychologist*, 1952, **7**, 340. (Abstract)

Journal of Consulting Psychology
Vol. 22, No. 3, 1958

The Relation of Difficulty in Drawing to Ratings of Adjustment Based on Human Figure Drawings

Marvin J. Feldman and Raymond G. Hunt
University of Buffalo

The clinical use of the human figure drawing proceeds from the assumption that a subject's style of expression reflects important facets of his personality. More specifically, it is assumed that those aspects of the self subsumed under the concept of "body image" will find direct expression in the drawing of the human figure. This point of view has been elucidated most fully by Machover (1). Her elaborations of the meaning of figure drawings seem to stem in part from empirical clinical findings. In large part, however, she implies that the phantasy life of the subject is communicated by a symbolic body language. Her interpretations often parallel psychoanalytic notions about symbolism, e.g., a concave mouth reveals concern with dependency needs.

Personality may be validly revealed in many aspects by the individual's style of expression. Nevertheless, the body language conveyed by figure drawings may not have such an exact syntax or logic as implied by Machover. Further, an interpreter may be led astray by influences other than symbolic representations. Whitmyre (3) has shown rather conclusively that the over-all aesthetic value of the drawing is highly related to clinical ratings of adjustment. In his study, ratings of personal adjustment correlated .88 and .86 with independent ratings of artistic excellence for the same drawings. It might be argued that any disturbance reflecting an inadequate body image would result in a departure from a smooth, regular drawing; hence, the over-all result would be displeasing. There is no good a priori reason for believing, however, that a self depreciatory body image would lead to irregularities in drawing in just those parts of the human figure which are most difficult to draw. The present study investigates possible structural factors in drawing apart from personality needs or traits which might unwittingly influence interpretations of figure drawings.

Procedure

As part of a larger project, 65 students, 39 men and 26 women, in a lower division course in psychology were asked to draw a human figure without clothes. The drawings were obtained in a regular classroom session. Although not specifically instructed to do so, all subjects drew their own sex. Three clinicians versed in the use of figure drawings independently rated 25 body parts on each drawing for specific signs of disturbance. The ratings were simple presence or absence of disturbance for each body part. To promote uniformity of ratings, the clinical judges were given brief descriptions of the signs of disturbance associated with each body part according to Machover (1), e.g., for nose—overly long or large, tiny "button nose," heavily shaded or cut off, heavy reinforcing of lines, nostrils present.

Two art instructors in figure drawing[1] who knew nothing about the study were asked to rate the same 25 body parts in terms of the ease or difficulty in learning to draw each body part. These ratings were done on a five-point scale in which 1 indicated the easiest and 5 the most difficult part to learn to draw.

This study has no direct bearing on the validity of the symbolic meaning of signs on

---

[1] The authors wish to thank Anne Garson and Alexander Russo of the Albright Art School of the University of Buffalo and Saul Siegel and Herbert Sanderson who served as judges in this study.

Table 1

Interjudge Reliability for Ratings of Disturbance

Judge	*r*
A vs. B	.77
A vs. C	.74
B vs. C	.86

figure drawings. It would seem to furnish some evidence of a possible source of error in figure drawing interpretation.

Results

The interrater reliability of the art instructors' ratings was only .50. Inspection of the data, however, shows that they disagreed with one another by more than one step on only three items. Hence, it seemed reasonable to combine their judgments into an average rating for each body part.

The interrater reliability of the clinical judges is shown in Table 1. These correlations indicate how well the judges agreed on which body parts (not which subjects) were most often rated as manifesting disturbance. It can be seen that the clinical judges agreed fairly well in checking signs of disturbance on the various body parts of the figure drawings. Inspection of the data, however, shows that one of the judges consistently checked more signs of disturbance in rating the drawings. Hence, instead of using a raw score average of the number of times each body part was checked, it seemed advisable to convert each clinical judge's ratings to a set of standard or *z* scores. The three sets of *z* scores were then averaged.

The main comparison was then made by correlating the average ratings of the two art instructors with the average *z* scores of the three clinical judges. The correlation was − .53, which is significantly different from zero at the .01 level. This negative correlation indicates that those body parts which are most difficult to draw tend also to be rated more frequently as manifesting signs of emotional disturbance.

Table *2* presents the various body parts ranked according to artists' ratings of drawing difficulty and frequency with which they are selected by clinicians as indicating emotional disturbance. According to the body-image theory underlying Machover's rationale, one would expect certain parts of the body to be selected by clinicians more frequently than others as reflecting emotional disturbance. It follows from Schilder's discussion (2) that those parts of the body associated with erotogenic zones (e.g., extensions, openings, genitalia, etc.) will be more highly cathected and hence more likely to provide a locus of disturbance in figure drawings. Reference to Table 2 reveals that this hypothesis is only partially fulfilled. Certain supposedly (theoretically) highly cathected body parts such as breasts, nose, eyes, etc. are conspicuous by their absence among the parts most frequently selected by clinicians.

Moreover, it will be noted that those body parts which do conform to expectancy by being frequently selected by clinicians tend

Table 2

Ranks of 25 Body Parts for Artists' Ratings of Difficulty of Drawing and for Frequency of Selection by Clinicians as Indicants of Emotional Disturbance

Body part	Disturbance frequency rank	Drawing difficulty rank
Hands	5	1.5
Fingers	2	3.5
Distribution of hair over head	20	7
Distribution of hair over body	16	13.5
Nose	10	13.5
Waist	19	23.5
Ears	24	13.5
Chin	18	13.5
Body build	12	20
Height	25	23.5
Trunk	15	13.5
Joints	21	7
Shoulders	9	13.5
Arms	3	7
Chest	17	20
Breasts	14	20
Hips	11	13.5
Legs	6	7
Feet	4	1.5
Posture	13	25
Weight	22.5	20
Sex organs	1	20
Mouth	7	3.5
Eyes	8	7
Eyebrows	22.5	13.5

also to be body parts rated by artists as difficult to draw (e.g., feet, hands, etc.). Similarly, those body parts which fail to conform to expectancy, by being relatively infrequently selected by clinicians, tend to be parts rated relatively easy to draw (e.g., nose, breasts, etc.).

The only important exception to this is the clinician's rating of sex organs as the most frequent source of signs of disturbance. Sex organs were rated as easy to draw. Thus, it seems that with respect to sex organs, clinician's judgments are determined by factors other than artistic excellence. On the other hand, in terms of body-image theory, it is difficult to account for the absence of such body parts as breasts and eyes from those most frequently selected by clinicians, except as a function of some "artistic" variable.

Discussion

As was indicated above, the present study makes no necessary presumptions concerning the validity of symbolic meanings of signs on the figure drawing. Rather, the focus is upon a possible source of error confounding clinical interpretations from such drawings.

It seems quite evident from the data obtained in the present investigation and from related studies by Whitmyre (3) and Woods and Cook (4) that a considerable portion of the variance in figure drawing interpretation may be attributed to structural aspects of the drawing as distinct from symbolic aspects presumed to reflect personality traits or dimensions. As Table 2 indicates, there is considerable overlap between those parts of the figure most frequently selected by clinicians as containing indications of emotional disturbance and those parts rated most difficult to draw by figure drawing instructors (artists). However, the fact that this overlap is something less than complete (e.g., sex organs frequently selected by clinicians but rated relatively easy to render by artists) suggests that factors other than drawing skill are also influencing the clinicians' judgment.

Yet the fact remains that clinicians are evidently strongly influenced by the "artistic" dimension, both in their over-all evaluations of the drawing and with regard to specific areas of the drawing upon which they base judgments. Consequently, the variance attributed to this source would have to be somehow separated out from other sources of variance if the figure drawing method is to qualify for serious consideration as an effective general diagnostic instrument.

Summary

In an attempt to investigate possible structural factors in drawing, apart from personality needs or traits which might unwittingly influence clinical interpretations of human figure drawings, unclothed human figure drawings were collected from 65 undergraduate college students. Three clinicians rated 25 body parts on each drawing for specific signs of emotional disturbance. Two art instructors then rated the same 25 body parts along a five-point scale of difficulty in drawing.

A correlation of −.53 was obtained between average ratings of clinicians and artists for the 25 body parts. This correlation indicates that those body parts which are most difficult to draw tend also to be selected more frequently by clinicians as manifesting signs of emotional disturbance.

The implications of this and other similar findings for the use of figure drawings in personality diagnosis are discussed.

Received June 21, 1957.

References

1. Machover, K. *Personality projection in the drawing of the human figure.* Springfield, Ill.: Charles C Thomas, 1948.
2. Schilder, P. *The image and appearance of the human body.* Psyche Monographs No. 4; London: Kegan Paul, Trench, Trubner, 1935.
3. Whitmyre, J. W. The significance of artistic excellence in the judgment of adjustment inferred from human figure drawings. *J. consult. Psychol.,* 1953, **17**, 421–424.
4. Woods, W. A., & Cook, W. E. Proficiency in drawing and placement of hands in drawings of the human figure. *J. consult. Psychol.,* 1954, **18**, 119–121.

Journal of Consulting Psychology
Vol. 22, No. 3, 1958

On the Relationship Between Manifest Anxiety and Intelligence[1]

Charles D. Spielberger
Duke University

The Taylor Manifest Anxiety Scale (A scale) has been utilized extensively as a means of selecting subjects who are assumed to differ systematically in motivational level (4, 5, 19). Subjects (*S*s) who scored high on the A scale have been found to perform relatively better on simple tasks (13, 20) and relatively more poorly on complex tasks (6, 13, 15, 21) than *S*s who scored low. However, negative correlations between A-scale scores and various measures of intelligence have been reported in several recent studies (1, 3, 11). Kerrick (9) found that the A scale was significantly negatively correlated with six different measures of intelligence or aptitude. Grice (8), moreover, has demonstrated that inferior performance on a complex task by high anxiety *S*s could be attributed to intellectual differences as readily as to differences in motivation. These findings have raised the question of the validity of the A scale as an *uncontaminated* measure of motivation.

On the other hand, Farber and Spence state that they have "been unable, over a period of years, to find any relation between the Anxiety Scale scores of college students and conventional measures of intellectual ability such as entrance-examination scores and grade point averages" (7, p. 10). Klugh and Bendig (10), Dana (2), Mayzner *et al.* (12), and Sarason (16) have also reported essentially zero correlations between the A scale and various measures of intelligence. Schulz and Calvin (17), who also failed to find a relationship between the A scale and intelligence, suggested that the contradictory findings cited above might be attributed to variations in selection procedures at different institutions. The present study is concerned with factors related to the selection of *S*s which may affect the relation between the A scale and intelligence in college students.

Method

The A scale was administered to students in the introductory psychology course at Duke University for six consecutive semesters from the fall of 1954 through the spring of 1957. A form of the test consisting of the 50 anxiety items plus the 15 L-scale items of the MMPI was utilized. McCreary and Bendig (14) have demonstrated that the A-scale items alone correlated .96 with the same items embedded in the 225-item "Biographical Inventory" employed by Taylor (18). The ACE Psychological Examination for College Freshmen (1949 edition) is routinely administered to all entering freshmen at Duke University as part of a battery of placement tests. Scores were available on both scales for 1142 *S*s. The data were analyzed separately for each sex for the total sample and for semester subsamples.

Results and Discussion

The means and *SD*s of the A-scale and ACE distributions and the Pearson product-moment correlations for the total sample and the various subsamples are summarized in Table 1.[2] The correlations between A scale and ACE for the total sample and for total males and total females were not significantly

---

[1] A paper based on part of the data of this investigation was presented at the meetings of the Midwestern Psychological Association in Chicago, May 3, 1957.

[2] The author acknowledges with appreciation the statistical and clerical assistance rendered by Ann L. Wescott and Sandra Eisdorfer.

Table 1

Means, Standard Deviations, and Correlations for A-Scale and ACE Distributions

Group	Males						Females					
	N	Mean ACE	ACE *SD*	Mean MAS	MAS *SD*	Pearson *r*	*N*	Mean ACE	ACE *SD*	Mean MAS	MAS *SD*	Pearson *r*
Spring, 1955	72	111.4	21.5	13.1	7.9	−.34**	18	128.7	21.4	10.7	5.5	.62*
Spring, 1956	79	116.8	21.1	11.3	6.3	−.21	44	124.5	16.9	11.9	7.1	−.02
Spring, 1957	101	120.5	20.9	15.5	7.2	−.16	38	134.3	17.8	14.4	6.2	−.17
Fall, 1954	140	120.3	32.6	12.4	7.8	−.13	119	128.2	18.3	13.4	5.9	.01
Fall, 1955	122	124.2	20.7	12.0	6.4	−.08	130	131.5	39.5	13.3	6.4	.09
Fall, 1956	160	124.6	18.2	14.8	7.0	.04	119	131.1	17.0	15.4	8.2	−.19*
Spring	252	116.8	21.9	13.6	7.2	−.19**	100	129.0	18.6	12.6	6.7	.08
Fall	422	123.0	24.6	13.2	7.2	.01	368	130.3	31.8	14.0	6.9	.00
Total	674	120.7	23.8	13.4	7.2	−.06	468	130.0	25.9	13.7	6.9	.01
Total sample	1142	124.5	25.1	13.5	7.1	−.02						

* $p < .05$.
** $p < .01$.

different from zero. Although two of the correlations for the females were significant, intensive analysis of the distributions of both variables for the individual and pooled samples failed to suggest any systematic interpretation of these findings.[3] Therefore, the remainder of this report is concerned only with male *S*s.

Correlations between A-scale and ACE scores for semester subsamples for the males ranged from − .34 to + .04. Only one of these *r*'s was significantly different from zero (two-tailed tests), but the magnitudes of the obtained correlations were found to be related to the mean ACE scores of the subsamples. The relationship between mean ACE and A-scale–ACE correlation is depicted in Fig. 1A. As the mean ACE increased, the extent of the negative correlation between the A scale and the ACE decreased. The rank difference correlation for this relationship was .94.

Since the *SD*s for the various subsamples of males were approximately equal, the fact that the negative correlations between the A scale and the ACE *decreased* and approached zero as the subsample ACE means *increased* suggested that the obtained correlations were due primarily to the presence of *S*s scoring in the lower ranges of the ACE distribution. The distributions of ACE scores for male semester subsamples were compared with the 1949 national ACE norms in Table 2. Chi-square tests of goodness of fit of the obtained distributions to the ACE norms were computed. Rank difference correlations between the percentage of *S*s with ACE scores below the 20th and 40th percentiles of the ACE norms in a given subsample and the A-scale–ACE correlation for that subsample were − .94 and − 1.00. There was also a perfect positive rank difference correlation between the magnitude of deviation of a given subsample from the ACE

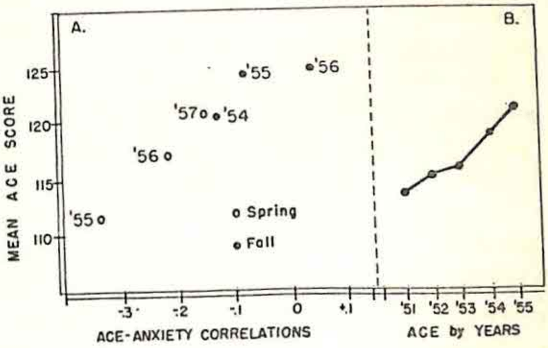


Fig. 1. *A*. Mean ACE scores and A-scale—ACE correlations for each of the six semester subsamples for males. *B*. Mean ACE scores for male freshmen by year of entry.

---

[3] The lower end of the range of ACE scores for the female sample was severely curtailed. Only .7% of the female *S*s scored below the 20th percentile of the national ACE norms and only 6.8% scored below the 40th percentile.

Table 2

Distribution of ACE Scores for Males in Semester Subsamples

Group	Below 88	89–102	103–114	115–128	Above 129	Chi square
ACE norms	20%	20%	20%	20%	20%	—
Spring, 1955	14%	25%	14%	24%	24%	4.5
Spring, 1956	9%	18%	20%	19%	34%	16.0*
Spring, 1957	4%	12%	24%	27%	33%	28.0**
Fall, 1954	7%	8%	21%	26%	38%	47.0**
Fall, 1955	3%	10%	14%	32%	41%	65.6**
Fall, 1956	2%	10%	15%	34%	39%	81.0**

* $p < .01$.
** $p < .001$.

norms as measured by chi square and the A-scale–ACE correlation.

Two types of selection factors seemed to operate simultaneously to affect the number of *S*s falling in the lower ranges of ACE scores for the male subsamples in this study. The trend of the ACE subsample means, which is depicted in Fig. 1A, indicated (*a*) that male *S*s who took the introductory psychology course in the fall tended to be brighter than those who took it in the spring (fall mean ACE 123.0; spring mean ACE 116.8; $F = 11.1$, $p < .001$), and (*b*) that mean ACE scores increased in successive spring subsamples and successive fall subsamples as a function of time. A number of poorly understood intra-institutional selection factors appeared to be responsible for the fact that brighter students tended to take the introductory course in the fall.[4] There were no differences, however, in either the means or the variances of the distributions of A-scale scores for fall and spring. Therefore, the differences *between* the fall and spring A-scale–ACE correlations seemed to result from the differences in the distributions of ACE scores.

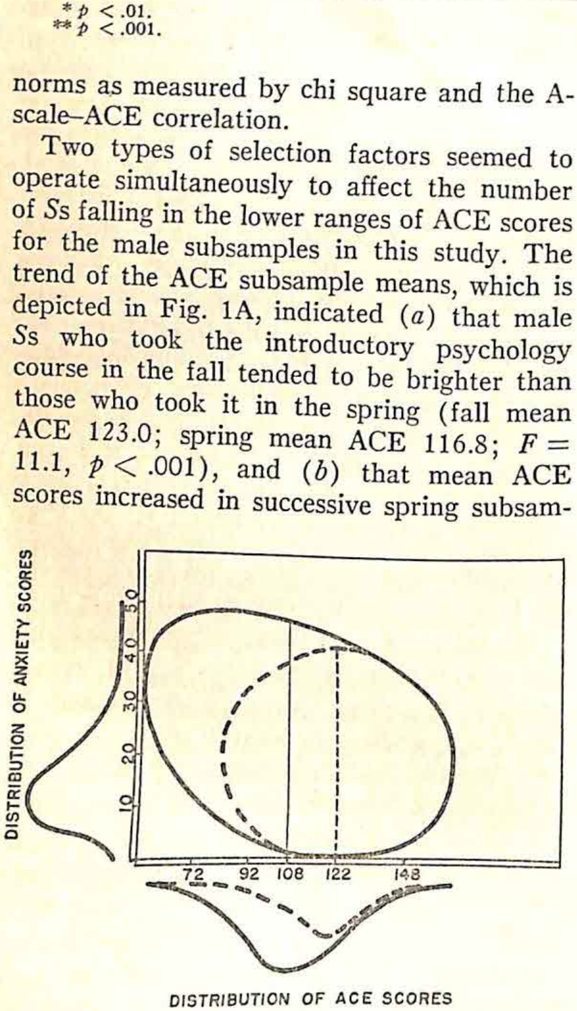


Fig. 2. An explanation of the effects of selection factors on the relationship between the A scale and the ACE. The solid lines represent the hypothetical sampling distributions of ACE and A-scale scores in an unselected population of college males. The dotted line represents the effects of selection factors which curtailed the range of the ACE distributions for this study (see Table 1 and Table 2).

The trends *within* the fall and the spring subsamples seemed to reflect the results of selection procedures which have systematically reduced the number of male students with low intelligence who have been admitted to Duke University. The mean ACE scores for entering male freshmen for the period of time affecting this study (fall, 1951—fall, 1955)[5] are shown in Fig. 1B. The trend of increasing mean intelligence in successive classes of entering freshmen over this five-year period was highly significant (22).

---

[4] The introductory psychology course at Duke University is a one-semester course, and staff members who have been associated with this course over a number of years have opined that in the spring this course attracts a greater number of students who have failed the first semester of a two-semester introductory course in another area. Also, the free elective system at the university is such that students who are out of step or more poorly organized are likely to take the introductory course in the spring.

[5] The data for entering freshmen were from those students who first entered Trinity College (male undergraduate arts and sciences) during the fall semester. Data were unavailable for male freshmen who entered during the spring and summer, but this amounted to only 7% of all Trinity College freshmen.

An hypothetical explanation of these data is depicted graphically in Fig. 2, in which a small negative correlation between the A scale and the ACE is posited for male *S*s. Where a sample contains a sizable number of cases in the lower ranges of ACE, the negative correlation between ACE and anxiety is more pronounced; where there are few cases with low ACE scores, the correlation approaches zero. Due to the operation of university selection procedures and intra-institutional selection factors in this study, the range of ACE scores is differentially curtailed at the lower end, resulting in the obtained correlations which varied as a function of the semester and year in which the introductory course was taken.

The extent to which this interpretation can be generalized to other populations is difficult to evaluate. Most studies of the relationship between anxiety and intelligence have not reported separate correlations for each sex and have not reported data on their intelligence score distributions. The above interpretation is consistent, however, with the significant negative correlations between the A scale and measures of intelligence obtained by both Grice (8) and Kerrick (9). These experimenters utilized male samples (airmen basic trainees) who would presumably be heterogeneous in intelligence with a sizable percentage of their *S*s falling in the lower intelligence range. Matarazzo and his associates (11) found a significant negative correlation between the A scale and the ACE ($r = -.25$) in a group of male college sophomores. This correlation appears to be somewhat inconsistent with the suggested interpretation of our findings in that the mean ACE score for Matarazzo's sample is very high (estimated by the present investigator from their Fig. 2 to be approximately 134; 87th percentile on 1949 ACE norms). Although data on the ACE distribution in the Matarazzo study is not specifically reported, the range of ACE scores appeared to be moderately heterogeneous and might include a sufficient number of cases with low ACE scores to account for the obtained correlation. In this same study, the authors failed to find any relationship between the A scale and either the CVS (an abbreviated version of the Wechsler Bellevue) or grade-point averages which points up the importance of the particular measure of intelligence used. Calvin *et al.* (1) found no relationship between A-scale scores and IQ as measured by the Wechsler Bellevue Intelligence Test in a group of college *S*s (21 males; 15 females) whose IQ range was 102–141 (mean 122.75, *SD* 9.18). When they added a second group of 15 *S*s (12 males; 3 females) with lower scores (range 92–129, mean 108.3, *SD* 10.41), the combined groups yielded highly significant negative correlations between the A scale and IQ. The emergence of significant negative correlations as a function of including a number of male *S*s with low intelligence is consistent with the interpretation of the data of the present study. Calvin *et al.* point out, however, that low IQ in their second group was confounded with scholastic difficulty in that *S*s in this group were obtained from a how-to-study course offered for students having scholastic difficulty.

While Kerrick's statement that it is "virtually impossible to select extreme subjects on the Taylor scale who are equated in intelligence" is obviously not consistent with the data, the implication of the studies cited above and the findings of the present study are that a small negative correlation between anxiety and intelligence may be found in male college *S*s if one samples a wide range of intelligence and if the sample contains a sizable proportion of *S*s from the lower part of the range. However, this correlation may be obscured by variations in selection procedures which curtail the range of intelligence. It should be noted that the present study offers no explanation for a relationship between the A scale and the ACE, nor does it answer questions pertaining to the validity of the A scale as a measure of motivation. It does suggest that experimenters selecting male *S*s on the basis of the A scale should control for intelligence, especially when studying heterogeneous populations with low mean intelligence.

Summary

Scores on the Taylor Manifest Anxiety Scale and the ACE Psychological Examination for College Freshmen were obtained for

six consecutive semesters from a total of 1142 students enrolled in introductory psychology courses. The correlations between the A scale and the ACE for the total sample, total males, and total females were not significantly different from zero. Although only one of six semester correlations for the males was significant, the magnitudes of these correlations were highly related to the mean ACE scores for these subsamples. Where a subsample contained a sizable proportion of *S*s with low ACE scores, the negative correlation between the A scale and the ACE was more pronounced; where there were few *S*s with low ACE scores, the correlation approached zero. It was suggested that a small negative correlation between the A scale and intelligence may be found in college males if one samples a wide range of intelligence and if the sample contains a number of *S*s from the lower part of the range. However, this correlation may be obscured by variations in selection procedures which curtail the range of intelligence.

Received June 18, 1957.

References

1. Calvin, A. D., Koons, P. B., Bingham, J. L., & Fink, H. H. A further investigation of the relationship between manifest anxiety and intelligence. *J. consult. Psychol.*, 1955, **19**, 280–282.
2. Dana, R. H. Manifest anxiety, intelligence, and psychopathology. *J. consult. Psychol.*, 1957, **21**, 38–40.
3. Duvall, Mary I. The effect of manifest anxiety upon incidental and new task learning. Unpublished doctoral dissertation, George Peabody College, 1956.
4. Farber, I. E. Anxiety as a drive state. In M. R. Jones (Ed.), *Nebraska symposium on motivation.* Lincoln: Nebraska Univer. Press, 1954.
5. Farber, I. E. The role of motivation in verbal learning and performance. *Psychol. Bull.*, 1955, **52**, 311–327.
6. Farber, I. E., & Spence, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120–125.
7. Farber, I. E., & Spence, K. W. Main and interactive effects of several variables on reaction time. *U. S. Naval Res. Lab.*, Mar. 1, 1955, Tech. Rep. No. 3. (Contract N9 onr-93802, State Univer. of Iowa.)
8. Grice, G. R. Discrimination reaction time as a function of anxiety and intelligence. *J. abnorm. soc. Psychol.*, 1955, **50**, 71–74.
9. Kerrick, Jean S. Some correlates of Taylor Manifest Anxiety Scale. *J. abnorm. soc. Psychol.*, 1955, **50**, 75–77.
10. Klugh, H. E., & Bendig, A. W. The Manifest Anxiety and ACE scales and college achievement. *J. consult. Psychol.*, 1955, **19**, 487.
11. Matarazzo, J. D., Ulett, G. A., Guze, S. B., & Saslow, G. The relationship between anxiety level and several measures of intelligence. *J. consult. Psychol.*, 1954, **18**, 201–205.
12. Mayzner, M. S., Sersen, E., & Tresselt, M. E. The Taylor Manifest Anxiety Scale and intelligence. *J. consult. Psychol.*, 1955, **19**, 401–404.
13. Montague, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91–96.
14. McCreary, J. B., & Bendig, A. W. Comparison of two forms of the Manifest Anxiety Scale. *J. consult. Psychol.*, 1954, **18**, 206.
15. Ramond, C. Anxiety and task as determiners of verbal performance. *J. exp. Psychol.*, 1953, **46**, 120–124.
16. Sarason, I. G. The relationship of anxiety and "lack of defensiveness" to intellectual performance. *J. consult. Psychol.*, 1956, **20**, 220–222.
17. Schulz, R. E., & Calvin, A. D. A failure to replicate the findings of a negative correlation between Manifest Anxiety and ACE scores. *J. consult. Psychol.*, 1955, **19**, 223–224.
18. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
19. Taylor, Janet A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53**, 303–320.
20. Taylor, Janet A., & Chapman, J. P. Paired-associate learning as related to anxiety. *Amer. J. Psychol.*, 1955, **68**, 671.
21. Taylor, Janet A., & Spence, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–64.
22. Weitz, H. Placement test performance of Duke University freshmen—a five year summary. Duke University, Bur. of Testing and Guidance, Staff Research Report No. 14, Nov. 15, 1955.

Journal of Consulting Psychology
Vol. 22, No. 3, 1958

Anxiety: Unitas or Multiplex?[1, 2]

Douglas N. Jackson
Pennsylvania State University

and Robert Bloomberg
VA Hospital, St. Cloud, Minnesota

One way of determining if a construct or dimension, such as anxiety, is unitary is to examine its relationships with other constructs and with the observables which help to define the constructs. If a construct is defined directly in terms of a number of independently measurable observables, one can look for empirical relationships among the observables which should reflect something about the construct. The conclusion that a single dimension is sufficient to account for all of the variables presumed to represent a construct like anxiety requires not only that all of the variables are substantially correlated, but also that the correlations among all of the measures are due to a single common factor to which all of the measures are mutually related. The present study seeks to assess relationships among a number of purported measures of anxiety in order to determine the degree to which these reflect a single dimension.

Method

Subjects

From the admissions to a neuropsychiatric hospital, 37 male patients were selected randomly, with the exceptions that those with known organic pathology, as well as those with ages in excess of 45 years, were excluded. The age range was 19–43 with a median of 30 years. Diagnostically, the group was composed as follows: schizophrenic reaction, 22; manic-depressive, 1; psychoneurotic (including depressive) reaction, 13; alcoholism, 1.

Procedure

Four measures were administered to each *S*: (*a*) The Wechsler-Bellevue digit span score, which Rapaport (10, pp. 185–193) has put forth as a measure of anxiety without confirmation in later studies (6), and a vocabulary scale (11) to permit statistical control of verbal intelligence in assessing digit span performance; (*b*) the Taylor Manifest Anxiety Scale (13), administered as part of the Minnesota Multiphasic Personality Inventory (MMPI); (*c*) two one-minute observations of the blink rate, recorded while the *S* viewed a drawing of a triangle mounted on a vertical plane 60 inches from his line of sight; (*d*) the fingerprint stain test of palmar sweating.

The blink rate. Meyer (7) reviews the empirical evidence for considering the blink rate as an indicant of anxiety. For example, when an individual is under stress, as he is while undergoing cross-examination in a court of law, the blink rate increases (9). Meyer offers a theoretical interpretation which attributes the interaction of simultaneous responses to a convergence of impulse patterns upon the motor pathways of the central nervous system. Variations in the blink rate are attributed to excitation from these neighbor-

---

[1] Portions of this paper were read at the meetings of the Midwestern Psychol. Ass., St. Louis, 1956. Most of this report was written while the first author was at the Menninger Foundation as a U. S. Public Health Service Clinical Research Fellow of the National Institute of Mental Health.

[2] Published with permission of the Chief Medical Director, Department of Medicine and Surgery, Veteran's Administration, who assumes no responsibility for the opinions expressed or conclusions drawn by the authors.

ing channels. The corrected split-half reliability of the blink rate index was .91 in the present study.

The fingerprint stain test of palmar sweating. Mowrer, Light, Luria, and Zeleny (8) review the successive developments of this device. It was validated by evidence that it decreases in therapy (1, 5) and in a relaxing situation (3), while increasing under stressful conditions. Anxiety neurotics have shown greater palmar sweating than "normals" (5).

The test, described by Mowrer *et al.* (8), consists of dabbing a solution of anhydrous ferrous chloride on the index finger and allowing it to dry for one minute. The *S* is then instructed to place his index finger (at one pound pressure as measured by a scale) on a piece of Dietzgen 198 M paper treated with a five per cent solution of tannic acid. The amount of ferrous chloride deposited on the paper, and hence the density of the fingerprint, has been shown to be a linear function of the amount of sweat (12). The density of the fingerprint is determined by the amount of light passing through the print by means of a specially constructed photoelectric circuit containing a microammeter. The repeat-reliability of this photometric reading of fingerprint density was in excess of .99 in the present study. The test-retest reliability of this measure of palmar sweating has been found to be .89 after a half-hour interval (14).

Results

Product-moment correlations were computed among the four measures. Table 1 indicates that none of these correlations are significantly different from zero. When the vocabulary score is partialled out of correlations involving the digit span measure, the range is from − .17 to + .02.

Table 1

Correlations Among Anxiety Measures

Measures	Digit span	Taylor Anxiety scale	Blink rate
Taylor Anxiety scale	.02*		
Blink rate	−.08*	−.11	
Palmar sweating	−.17*	−.15	−.02

* Covariance attributable to verbal knowledge partialled out.

Discussion

If these results concerning lack of relationship between assumed measures of "anxiety" prove reliable, and an extensive factor analytic study by Holtzman and Bitterman (4) supports this finding, a careful scrutiny of the conceptualization and experimental definition of anxiety seems in order.

One possibility is to consider anxiety as a second-order factor, and consider the primary factors to be such dimensions as muscular tension, palmar sweating, etc., analogous to the situation now prevailing in the area of human abilities. This would be true only if there was some correlation between the primary factors. The evidence on this point is incomplete, but if various proposed measures of anxiety can be placed along orthogonal axes and are completely unrelated, one no longer has a meaningful single dimension or grouping of correlated dimensions of anxiety, but is left with a less parsimonious, but possibly more realistic, structure of independent dimensions of such things as muscular tension, of palmar sweating, and, among other things, a dimension of test-taking behavior, the principal factor of which may be best exemplified by the Taylor scale or the MMPI *K* scale.

The evidence reported here cannot but raise questions concerning the generality of at least some of the measures employed. Widespread and uncritical use of the Taylor scale as an indicant of anxiety, for example, may lead to misleading results and erroneous conclusions based only to a minor extent upon variance associated with "manifest anxiety" and to a greater, if unknown, extent upon such determinants as response acquiescence (2), social desirability, or other superfluous factors.

Summary

In order to gather preliminary evidence concerning the extent to which four supposed indicants of anxiety could be considered to tap the same or related dimensions in a population of male functional psychiatric hospital patients, the following tests were administered: (*a*) digit span; (*b*) Taylor Anxiety scale; (*c*) fingerprint stain test of palmar sweating; and (*d*) blink rate.

Although the measures were acceptably reliable, their intercorrelations failed to depart significantly from zero. These results, when considered with other evidence, suggested that manifestations of anxiety are multidimensional rather than unitary and may even represent uncorrelated dimensions. This may necessitate a conceptual reinterpretation of anxiety and cautions against the uncritical acceptance of a single measure as an indicant of generalized anxiety.

Received June 10, 1957.

References

1. Bixenstine, V. E. A case study of the use of palmar sweating as a measure of psychological tension. *J. abnorm. soc. Psychol.,* 1955, **50**, 138–143.
2. Cronbach, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt,* 1950, **10**, 3–31.
3. Gladstone, R. An investigation of the relationship between palmar sweating and emotion as measured by a group test of palmar sweating. Unpublished doctoral dissertation, Univer. of Illinois, 1949.
4. Holtzman, W. H., & Bitterman, M. E. A factorial study of adjustment to stress. *J. abnorm. soc. Psychol.,* 1956, **52**, 179–185.
5. Light, B. H. Tension changes in patients undergoing psychotherapy. Unpublished doctoral dissertation, Univer. of Illinois, 1951.
6. Matarazzo, Ruth G. The relationship of manifest anxiety to Wechsler-Bellevue subtest performance. *J. consult. Psychol.,* 1955, **19**, 218.
7. Meyer, D. R. On the interaction of simultaneous responses. *Psychol. Bull.,* 1953, **50**, 204–220.
8. Mowrer, O. H., Light, B. H., Luriz, Z., & Zeleny, M. P. Tension changes in psychotherapy with special reference to resistance. In O. H. Mowrer (Ed.), *Psychotherapy: Theory and research.* New York: Ronald Press, 1953. Pp. 546–640.
9. Ponder, E., & Kennedy, W. P. On the act of blinking. *Quart. J. exp. Physiol.,* 1927, **18**, 89–110.
10. Rapaport, D. *Diagnostic psychological testing.* Chicago: Year Book Publishers, 1945.
11. Shipley, W. C. A self-administering scale for measuring intellectual impairment and deterioration. *J. Psychol.,* 1940, **9**, 371–377.
12. Silverman, J. J., & Powell, V. E. Studies on palmar sweating. *Psychosom. Med.,* 1944, **6**, 243–249.
13. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.,* 1953, **48**, 285–290.
14. Wenger, W. A., & Gilchrist, J. C. A comparison of two indices of palmar sweating. *J. exp. Psychol.,* 1948, **38**, 757–761.

Journal of Consulting Psychology
Vol. 22, No. 3, 1958

Test Anxiety Questionnaires[1]

George Mandler and Judith E. Cowen[2]
Harvard University

The increased use of paper-and-pencil tests for the measurement of anxiety makes it appropriate to communicate the following additional information on one of these instruments.

The college form of the Test Anxiety Questionnaire (TAQ) was first described by Mandler and Sarason (3). A high school form has since been developed by Cowen (1). These questionnaires were designed to elicit self-ratings on items descriptive of anxiety reactions in test situations. The college form contains 39 items, of which four are fillers; the high school form contains 52 items, including four fillers.[3] Each item consists of a graphic scale with the end points marked with written description and the midpoint indicated. The subject is free to mark each individual scale anywhere along the continuum.

Scoring

Originally, the TAQ was scored by dividing each scale into 10 intervals and assigning a score of 1 to 10 to the subject for each scale. A median was then established for the distribution on each individual scale and each subject received a score of zero or one depending on whether his scale response fell below or above the sample median. Sarason and Gordon (5) have presented scoring norms based on this system.

A preferable method of scoring the TAQ is to sum the raw scores on the individual items, thus eliminating the tedious determination of medians for each item, which may vary from sample to sample. An empirical comparison of the two scoring systems on two independent samples yielded product-moment correlations of .95 ($N = 81$) and .93 ($N = 41$). Thus, the new scoring system adequately fulfills the function of the old, while at the same time saving considerable time and effort and eliminating the need for norms specific to the sample.[4]

Reliability

The test-retest reliability of the college form of the TAQ is .91 ($N = 70$), the split-half reliability is .91 ($N = 100$) (6). The split-half reliability of the high school form of the TAQ is .90 ($N = 286$).

Norms and Sex Differences

The college form of the TAQ was administered to a group of Harvard-Radcliffe sophomores and juniors ($N = 99$) and a group of Boston University freshmen and sophomores ($N = 80$).[5] The high school form was administered to a group of Watertown, Massachusetts, high school sophomores ($N = 286$). Table 1 shows means and standard deviations for these groups by educational level and by sex.

All the means on the college form are near the midpoint of the theoretically possible

---

[1] Parts of this research were supported by a grant from the Laboratory of Social Relations, Harvard University.

[2] Now at Department of Education, Waterbury, Connecticut.

[3] Copies of the two questionnaires and detailed scoring instructions have been deposited with the American Documentation Institute. Order Document No. 5485, remitting $1.75 for microfilm or $2.50 for photocopies.

[4] If the TAQ is considered as a composite of 35 different scales, Gulliksen's discussion (2) of the determination of composite scores is relevant. The empirical relationship found here could be predicted from his analysis, namely, that "the test mean (has) no effect whatever on the relationship between a test and the composite of which it is a part."

[5] The data were supplied by Michael Maccoby.

Table 1

Means and Standard Deviations on Two Forms of the TAQ

Form and group	N	Mean	SD
College Form			
Harvard Radcliffe sample	99	179.8	38.6
Male subjects	66	179.8	38.5
Female subjects	33	179.7	39.6
Boston University sample	80	167.1	47.8
Male subjects	48	158.7	48.4
Female subjects	32	179.7	44.5
High School Form			
Watertown sample	286	245.9	41.1
Male subjects	151	239.17	41.5
Female subjects	135	253.55	39.2

range, which is 35 to 350. The Harvard-Radcliffe sample shows a higher mean than the Boston University sample. The only significant sex difference (women scoring higher than men) appears on the Boston University sample ($p = .05$). The Watertown sample on the high school form (which is scored on a nine-point scale) also has a mean near the theoretical midpoint (range is 48 to 432) and sex differences in the same direction as the Boston University sample, although these are not significant.

Test Anxiety and Other Measures

Since test anxiety has been hypothesized to be operative in test situations, the relation between the TAQ and intelligence tests is of interest. Sarason and Mandler (6) reported a correlation of $-.209$ ($p < .05$) between the TAQ and the Henmon-Nelson Test of Mental Ability, college form. Cowen (1) reported a similar correlation between the high school TAQ and the Otis Gamma of $-.185$ ($p < .02$).[6]

Finally, the relationship between the TAQ and the widely used Taylor scale of Manifest Anxiety (7) is of interest. Raphelson (4) reported a correlation of .53 ($N = 24$). A replication on a Harvard-Radcliffe sample resulted in a correlation of .59 ($N = 35$). This consistent positive relationship between the two scales indicates some communality of traits measured, but is too low to permit either substitution of the two scales for each other or any uncritical comparison of the results obtained by the use of these instruments.

Received May 23, 1957.

References

1. Cowen, Judith E. Test anxiety in high school students and its relationship to performance on group tests. Unpublished doctoral dissertation, School of Education, Harvard Univer., 1957.
2. Gulliksen, H. *Theory of mental tests.* New York: Wiley, 1950.
3. Mandler, G., & Sarason, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.,* 1952, **47,** 166–173.
4. Raphelson, A. C. The relationships among imaginative, direct verbal, and physiological measures of anxiety in an achievement situation. *J. abnorm. soc. Psychol.,* 1957, **54,** 13–18.
5. Sarason, S. B., & Gordon, E. M. The test anxiety questionnaire: Scoring norms. *J. abnorm. soc. Psychol.,* 1953, **48,** 447–448.
6. Sarason, S. B., & Mandler, G. Some correlates of test anxiety. *J. abnorm. soc. Psychol.,* 1952, **47,** 810–817.
7. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.,* 1953, **48,** 285–290.
8. Zweibelson, I. Test anxiety and intelligence test performance. *J. consult. Psychol.,* 1956, **20,** 479–481.

---

[6] Comparable correlations between a children's test anxiety questionnaire and intelligence test scores have been reported by Zweibelson (8).

Journal of Consulting Psychology
Vol. 22, No. 3, 1958

Allergy and the Need for Recognition

Norman S. Greenfield[1]
University of Wisconsin Medical School

The number of papers which have been devoted to discussions of possible relationships between emotions and allergy would seem to make any systematic review of the literature somewhat redundant. The interested reader will find reviews of this work in Leigh (5) and in Dunbar (1).

A central focus of discussion in recent years has been the formulation generally credited to French and Alexander (2) relating allergy to maternal rejection. Support for this finding, gained during the psychoanalysis of asthmatic patients was offered by Saul (10) who reported that unsatisfied longings for love contribute to allergic sensitivity. The numerous researches of Miller and Baruch (6, 7, 8) primarily utilizing material inferred from interviews and observations of play techniques, have been widely quoted and consistently indicate that ". . . maternal rejection appears to be an important factor in the personality and somatic symptom-patterns of the clinically allergic child" (6, p. 278).

More recently, Harris and Shure used the ". . . unhampered appraisal of the teacher who observes the child day after day in the natural setting of the classroom . . ." (3, p. 322) and, basing their conclusions on teachers' descriptions of childrens' behavior, reported that no discernible personality pattern differentiates asthmatic from nonasthmatic children. Additionally, these authors wisely pointed out that the range of clinical observational techniques which have been brought to bear on the attempted resolution of this problem are extremely susceptible to the error of subjective influence.

It was reasoned that if maternal rejection and a subsequent unsatisfied longing for love are indeed associated with allergic sensitivity, then a hypothesis derivable from a systematic extension of this classical formulation would be that the allergic individual would have a greater need for recognition than will individuals who do not demonstrate allergic sensitivity. This need for recognition might be thought of as an attempt to gain a kind of belated mastery over the impulses which have been frustrated during an earlier period. Thus the task of the present study was to subject to systematic experimental analysis the hypothesis that the need for recognition is greater in a group of individuals receiving treatment for allergy than in a control group.

Method

The Technique

The Iflund Test (4), which is based on the system of psychological needs contained in Murray's theory of personality (9), was administered individually to each of the subjects (*S*s). Essentially, this is a memory test which consists of 51 pictures, 17 for men, 17 for women, and 17 for both. Thus 34 pictures are used in any single test. Each of 28 pictures is designed to represent a specific need (e.g., autonomy, exposition, abasement, recognition, etc.) and they are presented serially at constant intervals to the *S*. Three "bland" pictures are shown at the beginning and three at the end of the series in order to minimize the effects of primacy and recency of exposure. Following each series of exposures the *S* is asked to describe those he remembers. The exposures are repeated until the *S* has recalled the entire series. Some are recalled after the first showing, others after several showings. The assumption is made that the

---

[1] The author wishes to thank Seymour B. Crepea, Associate Professor of Medicine, for his technical assistance and helpful comments during the course of this research.

order in which the *S* recalls the pictures is related to his personality organization. Additionally, representations of needs which are being gratified are recalled earlier, and representations of ungratified needs are recalled later. It was hypothesized that since the need for recognition of the allergic individual is manifest rather than repressed, the pictorial representation of this need would be recalled earlier. Previous investigation (4) clearly demonstrated that the order of recall is determined primarily by the meaning of the pictures which is in turn related to the personality of the individual.

In the present study, the actual testing was done by a psychologically unsophisticated research assistant who was not aware of the specific hypothesis under examination. It was felt that this would serve as further insurance against the possibility of the examiner's communicating subliminal cues to the *S*.

Subjects

The experimental group was composed of 35 students, 20 males and 15 females, who were receiving regular immunologic treatment for allergic rhinitis and/or asthma at the allergy clinic of the University Student Health Department. A purely operational definition of allergy in terms of patients who had been placed on an immunologic regimen by either their own physicians or members of the clinic staff was employed. In short, allergy was defined only in terms of those individuals who were receiving treatment for allergy. No effort was made to recheck or validate the diagnosis. Under these relatively loosely defined conditions it is quite likely that some misdiagnosed individuals were included in the experimental group. However, it can readily be seen that to the extent to which this is so, the resultant error would tend to attenuate rather than confirm the hypothesis.

The technique of matched control was used with each allergic patient being paired with a nonallergic peer for sex, age, and level of education. The mean age of the group was 20.97 years with a range from 17 to 29 years. Controls were drawn from various extracurricular college groups.

Results

In order to establish a quantitative measure or a recall score, a *T* score was computed for each need picture expressing the number of showings for that picture relative to the total number of showings required. This method tends to equate the *S*s in terms of general memory ability since the actual score is a ratio. The mean *T* score for both groups was 70.07. The mean *T* score for the allergy group was 62.2, while that for the control groups was 77.9. The nonparametric sign test was employed in the analysis of the data, and this difference was found to be significant at beyond the .01 level of confidence, demonstrating that the ideational representation of the need for recognition was recalled significantly earlier among the allergic group than among the controls. No differences between the sexes were observed nor were the groups differentiated by any of the other variables which made up the total test.

Discussion

These results suggest confirmation of the thesis that, within the definition of the samples employed, the need for recognition is prominently associated with the personalities of patients who are on an immunologic regimen. The fact that the significant difference occurred in the form of early recall suggests that this need is, at some level, being satisfied, possibly by the ritual which is associated with allergy.

It should be made abundantly clear that there is no intent to imply that these data denote a causal relationship. The present study was designed to provide a methodologically rigorous test of an essentially clinical assumption, and it is felt that this task has been fulfilled. Whether the constellation of factors which are inferable from these findings are cause for, or a reflection of, the allergic condition remains a problem for future research in this area.

Summary and Conclusion

The hypothesis that there is a greater need for recognition among a group of individuals being treated for allergic sensitivity than among a control group was tested with a university student health population. Within the definition of the method and the sample the hypothesis was confirmed. It was emphasized that no causal relationship was implied by these findings.

Received July 1, 1957.

References

1. Dunbar, H. Flanders. *Emotions and bodily changes.* (4th ed.) New York: Columbia Univer. Press, 1954.
2. French, T. M., & Alexander, F. Psychogenic factors in bronchial asthma (Part I). *Psychosom. Med. Monogr.*, 1941, No. 4.
3. Harris, M. C., & Shure, N. A study of behavior patterns in asthmatic children. *J. Allergy*, 1956, **27**, 312–323.
4. Iflund, B. Selective recall of meaningful material as relative to psychoanalytic formulations in certain psychiatric syndromes. Unpublished doctoral dissertation, Univer. of California, 1953.
5. Leigh, D. Asthma and the psychiatrist; a critical review. *Internat. Arch. Allergy*, 1953, **4**, 227–246.
6. Miller, H., & Baruch, Dorothy W. Psychosomatic studies of children with allergic manifestations. *Psychosom. Med.*, 1948, **10**, 275–278.
7. Miller, H., & Baruch, Dorothy W. Marital adjustment in the parents of allergic children. *Ann. Allergy*, 1950, **8**, 754–760.
8. Miller, H., & Baruch, Dorothy W. A study of hostility in allergic children. *Amer. J. Orthopsychiat.*, 1950, **20**, 506–519.
9. Murray, H. A. *Explorations in personality.* New York: Oxford Univer. Press, 1948.
10. Saul, L. J. Some observations on the relation of emotion and allergy. *Psychosom. Med.*, 1941, **3**, 66–71.

Journal of Consulting Psychology
Vol. 22, No. 3, 1958

Chance Frequency in Matching Problems When Success or Failure is Reported After Each Matching Operation

Alan H. Roberts
Elgin State Hospital

Despite the relative importance of matching procedures in clinical investigations, the sources most often consulted by psychologists contain very little information concerning the statistical treatment of these kinds of data. Among the exceptions are an article by Dudek (1) which provides a formula and tables for determining chance probabilities when *n* different categories are matched with *n* different objects, and a discussion by Mosteller and Bush (2) concerning the statistical analyses of several matching designs.

An extremely useful matching procedure whose treatment has been neglected is one in which *n* different objects are to be matched with *n* different categories and the judge is to be informed of the success or failure of each matching operation as he proceeds. An example of this procedure might be an experiment in which clinical psychologists are

Table 1

Chance Frequencies (Roman Type) and Proportions (Bold-faced Type) When *n* Items are Distributed into *n* Categories and Success or Failure is Reported After Each Match

Items (*n*)	Number of correct matches (*r*)										Total orders (*n*!)
	1	2	3	4	5	6	7	8	9	10	
2	1 **.5000**	1 **.5000**									2
3	2 **.3333**	3 **.5000**	1 **.1667**								6
4	6 **.2500**	11 **.4583**	6 **.2500**	1 **.0417**							24
5	24 **.2000**	50 **.4167**	35 **.2917**	10 **.0833**	1 **.0083**						120
6	120 **.1667**	274 **.3806**	225 **.3125**	85 **.1181**	15 **.0208**	1 **.0014**					720
7	720 **.1429**	1764 **.3500**	1624 **.3222**	735 **.1458**	175 **.0347**	21 **.0042**	1 **.00020**				5040
8	5040 **.1250**	13068 **.3241**	13132 **.3257**	6769 **.1679**	1960 **.0486**	322 **.0080**	28 **.00069**	1 **.000025**			40320
9	40320 **.1111**	109584 **.3020**	118124 **.3255**	67284 **.1854**	22449 **.0618**	4536 **.0125**	546 **.0015**	36 **.000099**	1 **.000003**		362880
10	362880 **.1000**	1026576 **.2829**	1172700 **.3232**	723680 **.1994**	269325 **.0742**	63273 **.0174**	9450 **.0026**	870 **.00024**	45 **.000012**	1 **.0000003**	3628800

asked to match n Rorschach protocols with n different diagnostic categories. In such an experiment, the *subjective* probabilities in favor of the judge might be increased if he were told the success or failure of each match as he proceeds so that he might be given a second chance to match a given protocol if the first match were in error. The judge matches the first protocol with one of the diagnostic categories. If the match is correct, he is so told, and the protocol together with its diagnostic category is removed. If the match is incorrect, he is so told, and the protocol together with its *correct* diagnosis is removed. Either correct protocols or correct diagnoses may be removed so long as the procedure first used is continued throughout the experiment. The procedure is repeated until n matches have been made. The experimenter must then determine if the number of correct matches achieved is likely to have arisen by chance.

When a procedure of this type is followed, the tables supplied by Dudek are not applicable. The correct chance frequencies and their corresponding proportions are shown in Table 1 for $n = 2$ to 10. Where $n > 10$, the correct frequencies can be derived from this table by means of the formula

$$X_r^n = (n-1)(X_r^{n-1}) + X_{r-1}^{n-1}$$

where X_r^n is the chance frequency of obtaining r correct matches out of n pairs.

Received June 10, 1957.

References

1. Dudek, F. J. Determining "chance success" when a specific number of items are sorted into discrete categories. *J. consult. Psychol.*, 1952, **16**, 251–256.
2. Mosteller, F., & Bush, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. 1. *Theory and method*. Cambridge: Addison-Wesley, 1954. Pp. 289–334.

Journal of Consulting Psychology
Vol. 22, No. 3, 1958

An Unprofitable Application of the Semantic Differential[1, 2]

Jack Block

Institute of Personality Assessment and Research, University of California, Berkeley

Since the development of the semantic differential (2), the technique has achieved a variety of research applications. Central to a number of its usages has been the assumption that in some degree, in some fashion, the technique reveals relationships among concepts of which the subject may be unaware. If not yet deemed a measure reflecting of the unconscious, a number of investigators still have suggested the technique "pulls" information related to some underlying, complexly symbolic mediating process which is a powerful determinant of behavior but which the subject is unable to report directly.

This assumption is an important one to test, for, if it is correct, we can bring the technique to bear upon a large number of substantive problems. If the assumption is in fact unwarranted, then we are engaging in an uneconomical self-deception where more direct and simpler procedures are already available. The present note reports on an empirical test of the assumption of the "depth-penetration" of the semantic differential.

A sample of 112 students in a large introductory course in psychology at the University of California served as subjects (*S*s). Each *S* was provided a booklet and asked to describe his ideal self and his like-sexed parent by means of the semantic differential technique. The connotative scales employed for these descriptions came from an Osgood report (2), and a seven-point rating continuum was specified.

Following the semantic differential descriptions, each *S* was asked to check a list of 80 adjectives so as to describe his ideal self and then to check the adjective list so as to portray his like-sexed parent. Adjectives were to be identified as "characteristic or true," as "uncharacteristic or false," or as not especially relevant for the description.

Where the semantic differential descriptions were recorded on separate pages of the test booklet, the adjective descriptions were recorded on immediately adjacent (and immediately comparable) columns. This arrangement of material was intended to reduce the probability of direct comparison by the *S* of his semantic differential descriptions while accentuating the likelihood of comparison with the adjective list.

From these data, two sets of scores were derived. The similarity of one's ideal self to one's parent represents a kind of identification (4, 5). The similarity of the two semantic differential descriptions provides one measure of this identification and was quantified by computing the "distance" (1, 3) between the semantic descriptions. Presumably, the extent of identification as measured in this manner was *not* under voluntary control by the *S*s. The two adjective descriptions were similarly contrasted and a "distance" score computed for each comparison. Because of the apposition and manifestness of the adjective characterizations, it may be presumed that identification as measured in this way is under appreciable control by the participating *S*s.

The anticipation with which this study began was that some individuals would show a closeness of ideal self and like-sexed parent descriptions as measured by the direct adjec-

---

[1] This investigation, part of a larger research project, was supported by Research Grant M-1078 from the National Institute of Mental Health of the United States Public Health Service.

[2] The writer is indebted to John Enright for his help in evaluating the results of this study.

tive check-list approach but marked dissimilarity of ideal self and parent as measured by the semantic differential. It was hypothesized that *S*s with this pattern of identification would prove to be unusually suppressive or repressive individuals, the latter characteristics being measured independently.

The intended analysis never proved worthwhile, for when the two distributions of discrepancy scores were correlated, the *r* obtained, based upon an *N* of 92, proved to be .73. This correlation, when corrected for attenuation became .94. A corrected correlation of this magnitude means that the disguised and manifest means of evaluating identification were functionally equivalent in the way they ordered individuals. Information accessible by the one approach was completely accessible also by the other means.

For the writer, this finding cannot be ascribed to peculiarities of the present sample because it seems likely that in a college sample of this size, a sufficient number of repressive and misidentified *S*s exists to introduce discrepancy between the check list and differential indices.

Some details of the data analysis should be spelled out. The group of 112 *S*s was partitioned into four subgroups on the basis of different kinds of response sets observed in the data (e.g., dichotomous rather than trichotomous response to the adjective check list, use of a six-point instead of a seven-point continuum in rating the semantic differential scales, etc.). Analysis of variance indicated three of these subgroups fairly could be combined and the result cited above is based upon this larger sample because of the greater stability afforded the primary result.

Each of the subgroups was analyzed separately, with respect to the correlation of the two measures of identification. The correlations between the semantic differential and adjective check-list measures, corrected for attenuation, ranged from .67 to 1.20 (this latter excess is presumably due to a fortuitously low estimate of the reliability of the semantic differential measure), so that the present findings appear to be generalizable over the various response sets observed. In the four independent comparisons of the subgroups, the split-half reliability of the semantic differential identification measure was always less than that of the adjective check-list measure. For the *N* of 92, the semantic differential and adjective check-list split-half reliabilities are .70 and .86, respectively.

Although the semantic differential justifiably will continue to enjoy widespread usage, especially with reference to problems of meaning and phenomenology, it would appear from the results of the present study that the technique has certain limitations when it is used to develop *derived* measures of psychological functions. In this kind of research context, the semantic differential may be a rather complicated way of developing a measure more readily and reliably secured by other means.

Received July 3, 1957.

References

1. Cronbach, L. J., & Gleser, Goldine C. Assessing similarity between profiles. *Psychol. Bull.*, 1953, **50**, 456–473.
2. Osgood, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, **49**, 192–237.
3. Osgood, C. E., & Suci, G. J. A measure of relation determined by both mean difference and profile information. *Psychol. Bull.*, 1952, **49**, 251–262.
4. Sanford, N. The dynamics of identification. *Psychol. Rev.*, 1955, 62, 106–118.
5. Stoke, S. M. An inquiry into the concept of identification. *J. gen. Psychol.*, 1950, **76**, 163–189.

Journal of Consulting Psychology
Vol. 22, No. 3, 1958

Varieties of Response Sets[1]

William E. Broen, Jr.
University of Kansas

and Robert D. Wirt
University of Minnesota

The importance of response sets has long been recognized in animal experimentation. There has recently been an increasing awareness of the usefulness of measures of response sets in psychodiagnostic work. Whether response set scales are seen as possible potential measures of valid variance as Cronbach would think (1, 2), or as measures of something which must be corrected for or suppressed as Fricke suggests (4), all authors seem to agree that understanding of response sets and the development of scales to measure them are necessary for maximal validity in psychological testing. The primary need in this area seems to be more knowledge of the interrelations of response sets.

Although it has been assumed that there are many types of response sets and many measures of supposedly different response sets have been constructed, there is still no accurate empirical knowledge as to whether or not all of these scales measure the same broad general response set, or whether there are a few specific response sets or many unrelated response sets.

Method

It was believed that the above questions could best be answered by factor analysis. Because the completeness of the factor-analytic results would depend upon the completeness of the coverage of the response-set area, the literature was first searched for previously hypothesized varieties of response sets. Measures for 11 possible response sets were then developed; these were: tendency to agree, tendency to disagree, evasiveness, tendency to go to the extreme in agreeing or disagreeing, carefulness, speed, tendency to guess, tendency to check many items, tendency to list many items, tendency to spread judgments on a rating scale, and tendency to rate to the left or right of center. A previously studied set to favor certain answer positions in multiple choice situations was not included because of evidence which cast doubts on existence of such a set (2). Response-set consistency (i.e., stability of set demonstrated by testing with another scale) was also measured in the present study for the tendencies to agree, disagree, and evade, and for the tendency to spread judgments on a rating scale.

An attempt was made to keep the possible influence of item content on the test responses at a minimum.[2] This was done in part by minimizing content, e.g., part of the rating scale data was gathered under directions stating that it was a study in ESP and the subjects were to try to guess the pattern of a person who had made ratings three months previously. In those sections where content was necessary, an effort was made to minimize its influence by using "neutral" content as much as possible. For instance, the content for

---

[1] We wish to express our gratitude to Sarah Holbrook, senior psychologist, Minneapolis Public Schools; Walter Chapman, principal of Marshall High School, Minneapolis; Frances Anderson, teacher of senior problems at Marshall High School and her students whose generous cooperation made possible the collection of data reported in this study. We also wish to acknowledge the assistance of Chú Cheng-Ping in the analysis of the data.

[2] A copy of the Questionnaire has been deposited with the American Documentation Institute. Order Document No. 5539, remitting $1.75 for 35 mm. microfilm or $2.50 for 6 by 8 in. photocopies.

measures of tendency to agree, disagree, and evade was Fricke's Set *T*, which consists of those items of his Opinion Attitude and Interest Survey (5), which were marked as true by 40 to 60 per cent of each of his criterion groups.

The scales were then administered to three complete classes of high school seniors, all about 17 years old. Fifteen response-set scores (11 response sets, four reliability scores) were obtained for each of the 69 subjects.

Each of the 15 response-set variables was then correlated with each of the others (product-moment correlations) and the resulting matrix was factor-analyzed by the centroid method. In order to achieve more nearly accurate final loadings, the extraction of factors was continued to one factor beyond those indicated to be significant by the application of Humphries' Rule and Tucker's Phi, and all factors were carried in rotation. The nonsignificant factor was "residualized."

Results

Three of the four response sets checked for reliability showed consistency in retesting with different scales. The two measures of evasiveness correlated + .66; those for tendency to agree correlated + .56; and those for tendency to disagree correlated + .58. These correlations are significant beyond the .01 level. There was no intertest consistency for the tendency to spread ratings on a rating scale. The correlation between two measures of this previously hypothesized response set was + .03. A fifth pair of measures which could be considered as a measure of response-set consistency were the variables: tendency to check many adjectives as descriptive of the self, and tendency to list many adjectives in describing the self. These two variables were also significantly correlated ($r = + .41$).

Additional results included the finding that the tendency to go to the extreme in agreeing or disagreeing with personality statements about the self was correlated significantly with both measures of acquiescence (+ .41 and + .45) but was not correlated with the measures of tendency to disagree and that speed was not significantly correlated with carelessness. In fact the correlation, although insignificant (+ .13), tended to show that those who were most careless tended to work more slowly.

Table 1

Rotated Factor Loadings

Response set	Factor I	Factor II	Factor III	h^2
1. Tendency to go to extremes on "neutral" questions	.61	.11	−.07	.39
2. Tendency to agree with "neutral" items	.85	−.25	.03	.79
3. Tendency to disagree with "neutral" items	−.46	.71	−.07	.72
4. Tendency to be undecided on "neutral" items	−.65	−.51	.02	.68
5. Tendency to agree with "neutral" items	.71	.03	−.01	.51
6. Tendency to disagree with "neutral" items	−.31	.82	.00	.77
7. Tendency to be undecided on "neutral" items	−.33	−.78	.02	.72
8. Tendency to guess (small score = guessing)	.13	−.12	.03	.03
9. Tendency to carelessness (high score = carelessness)	.07	−.19	−.09	.05
10. (Time) Tendency to work quickly (high score = more time)	−.07	.34	−.15	.14
11. Tendency to check many adjectives descriptive of self	−.02	−.10	.53	.29
12. Tendency to spread ratings	.05	.01	−.07	.01
13. Tendency to list many adjectives descriptive of self	−.18	−.03	.65	.46
14. Tendency to spread ratings	.13	.03	.36	.15
15. Tendency to rate to the left of center on rating scales or, tending to rate "good"	−.05	.03	.16	.03

Application of Humphries' Rule and Tucker's Phi indicated that three factors were necessary to account for the significant portion of the common variance. The intercorrelations of the factors are as follows: Factors I and II, − .16; I and III, + .01; II and III, .00. The loadings for the 15 response-set variables on these factors are listed in Table 1.

Discussion

The results of this study corroborate those of other studies which show that test results are determined in part by tendencies to respond to items in certain ways regardless of the specific content of those items. Response-set stability during two testings with different scales was demonstrated for the tendencies to agree, disagree, evade, and the tendency to list or check many items in an adjective check list. The data contraindicated the existence of a consistent response set of spreading vs. narrowing of ratings on rating scales or a speed vs. carefulness set.

A study of the factor loadings for Factor I indicates that it is a dimension with tendency to agree that a statement is true, to make a decision and state it on the one hand; versus the tendencies to deny, to evade, or to be undecided on the other.

The second factor seems to be a dimension of disagreeing with assertions vs. evading the question or being undecided. These data indicate that tendency to disagree is not simply the opposite of an acquiescence set and both cannot be measured on a single scale. Tendency to disagree can be either a tendency to evade the committing of oneself to a particular assertion, and thus much like indecision, and opposed to the making of a positive statement; or it can be the making of a definite negative assertion, and thus be the opposite of evasion or indecision. The first factor can then be seen as a committing onself vs. evasiveness factor (note that tendency to go to the extremes—either very true or very false—has a loading of + .61), and the second factor can be seen as a positive denial vs. indecision factor. It is to be noted that although two scales are needed to measure agreement and denial, these two scales in combination define all points on the plane in which tendency to go to the extreme and indecision tendency are also located.

Factor III seems to indicate a rather specific tendency to list or check many adjectives.

This three-factor analysis does not, of course, indicate that all response sets can be measured by three scales. Rather, one of its main conclusions is the specificity of response sets. Two scales can be used to measure what was previously thought of as the separate denial, acquiescence, tendency to go to the extremes, and evasiveness sets, and a third can measure reasonably well both the hypothesized separate sets of number of adjectives listed and number checked. However, the low commonality on the other response sets shows a need for further scales to measure them. Other studies have reported consistent sets to work for speed, to guess, and to place check marks in certain portions of rating scales (see Cronbach's review of the literature [2]). But each of these demonstrated response sets received very low loadings on each of the factors isolated in our study and each had low intercorrelation with all the other response-set measures. Additional scales would be needed to measure them.

There is a need for measures of, and further research on, all the varieties of response sets. The feeling that all that is necessary is the study of only those response sets which can enter into the particular type of scale used in any particular diagnostic setting stems from the common, but erroneous, view that response sets are only something to be avoided or corrected away. For instance, Fricke (4) stresses only the usefulness of measures of response sets as suppressor scales and goes on to say "Optimal weighting of suppressor and uncorrected scores should probably result in *no correlation between the suppressor and corrected scores*" (the italics are Fricke's [4], p. 163). Statements such as the above are accurate only in those cases where there is no correlation between the response set itself and the criterion for which prediction is being attempted. If there is a correlation between a response set and a criterion, then either the "suppression" of that set *or the neglect to use tests in which that response set can operate* may lead to lower validities.

The clinical meaning of such response sets is not altogether clear. However, other work (5) has indicated, as we would emphasize, that these sets are extremely important in psychodiagnostics. We might speculate, for example, that our Factor III is related to energy level or to extroversion, and personality scales measuring these dimensions should then account for such a factor. Or, it is possible that there are development trends in response sets which, as we suggested in an earlier study (6), distort the results of personality questionnaires. Our Factor III, for example, may primarily reflect a phenomenon of adolescence. The same may be said for other factors. In an unpublished study by the authors, using college sophomores, we found, for example, high loadings on a common factor for: tendency to say "true" to MMPI items, the Taylor Manifest Anxiety Scale score, the MMPI scale *F* (all high negative loadings) with MMPI scales *L*, *Hy*-subtle, and *K* (all high positive loadings). Looked at another way, 50% of the variance of the *A* scale can be accounted for by certain "test-taking attitude" scales. We believe these relate to response sets and are operative whether the material is relatively "neutral" or "meaningful" clinically. It would seem to us naive to think that clinically important personality syndromes would not include certain response sets as relevant *clinical* features. For these reasons, further research in this area is needed to find out what are the relations of response sets to the various criteria in which we are interested.

Summary

The consistency and interrelations of 11 hypothesized varieties of response sets were studied. A factor analysis of the intercorrelation of these response sets yielded three factors: a committing oneself vs. an evasiveness factor, a positive denial vs. an indecision factor, and a tendency to list or check many adjectives factor. The existence of clinically relevant response-set factors was hypothesized. The uses of measures of the varieties of response sets and the need for further research with them were discussed.

Received July 10, 1957.

References

1. Cronbach, L. J. Response sets and test validity. *Educ. psychol. Measmt*, 1946, **6**, 475–494.
2. Cronbach, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt*, 1950, **10**, 3–31.
3. Fricke, B. G. *The Opinion Attitude and Interest Survey*. Minneapolis: Investor Diversified Services, 1955.
4. Fricke, B. G. Response set as a suppressor variable in the OAIS and the MMPI. *J. consult. Psychol.*, 1956, **20**, 161–169.
5. Meehl, P. E., & Hathaway, S. R. The K factor as a suppressor variable in the MMPI. *J. appl. Psychol.*, 1946, **30**, 525–564.
6. Wirt, R. D., & Broen, W. E., Jr. The relation of the Children's Manifest Anxiety Scale to the concept of anxiety as used in the clinic. *J. consult. Psychol.*, 1956, **20**, 482.

PSYCHOLOGICAL TEST REVIEWS

New Tests

Berdie, Ralph F., & Layton, Wilbur R. *Minnesota Counseling Inventory*. Grades 9–12. 1 form. Untimed, (50) min. IBM, Testscor, or hand scoring. Inventory booklet ($3.50 per 25); IBM answer sheet and profile sheet ($3.60 per 50); Hankes answer sheet ($1.90 per 50); profile sheet only ($1.90 per 50); manual, pp. 27, with hand-scoring keys (60¢), with machine-scoring keys (70¢); specimen set (75¢). New York: Psychological Corp., 1953, 1957.

The Minnesota Counseling Inventory is an instrument for counseling high school students, based on two earlier questionnaires. From the Minnesota Personality Scale (1941), three scales were taken: Family Relationships (*FR*), Social Relationships (*SR*), and Emotional Stability (*ES*). Five MMPI scales represented by selected and sometimes edited items are *L*, *Pd*, *Sc*, *D*, and *Si*, which in the new Inventory are designated, respectively, as Validity (*V*), Conformity (*C*), Adjustment to Reality (*R*), Mood (*M*), and Leadership (*L*). Forty-four of the 355 items are not presently scored but represent potentialities for the development of new scales.

The extensive data given in the Manual reveal both favorable and unfavorable features of the development of the Inventory. Both split-half and retest reliabilities are mainly satisfactory, ranging from .56 to .95 in two-grade, one-sex groups, with the typical coefficient in the middle .80's. Standard score norms, given separately for boys and girls and for grades 9–10 and 11–12, are based on adequate methods and numbers. The manual's suggestions for interpretation are explicit and appropriately modest.

One weakness is revealed by the data on validities and intercorrelations. Validities of entire scales are reported in terms of the CRs of differences between the scores of random samples of pupils and of extreme cases nominated by teachers and school administrators. There is no indication of item analysis against a criterion. Some scales, such as *SR*, differentiate well, but others, such as *R* (*Sc*), fail to do so. The tables of intercorrelations show that most *r*'s run from .35 to .60, but that the correlations between *SR* and *L*, and between *ES* and *R*, are as high as the scales' reliabilities. The authors might well have refined the scales further by item analysis, or at least might have warned users that the evidence points to five interpretable scores, not seven. Even so, the Inventory is a serviceable instrument for counselors whose knowledge of test construction permits them to be aware of its limitations.—*L. F. S.*

Lamke, Tom A., & Nelson, M. J. *Henmon-Nelson Tests of Mental Ability*. (Rev. ed.) Grades 3–12. 3 levels, 2 forms at each level. 30 (40) min. Test booklet, each level ($3.00 per 35), with manual, pp. 24, and record sheet; specimen set (42¢). Boston: Houghton Mifflin, 1957.

The widely used Henmon-Nelson, which first appeared in 1931, has been given a thorough revision and restandardization. The test is now published in three levels, for grades 3–6, 6–9, and 9–12. Each level contains 90 multiple-choice items in spiral-omnibus form, with several types of verbal items predominating; there are some numerical items and figure analogies. A fourth level, for college students, is in preparation. The test booklets are consumable and are scored rapidly by the use of a carbon-paper backing that eliminates the need for a key. Norms for each level are provided in terms of percentile ranks in grade, grade equivalents, MAs and deviation IQs.

The content of the revision is based on an extensive item analysis of the older tests, and the equivalence of the Forms A and B is well established. The separate manuals for the three levels give impressively favorable data about reliability, congruent validity with other tests, and concurrent and predictive validity with achievement measures. Reliabilities generally exceed .90, correlations with other intelligence tests approximate .75, and correlations with achievement are high, typically .70 with relevant tests and .60 with school grades. The standardization is a model of precise method, using a predetermined sample of 250 classrooms stratified by state and size of school system. The manual closes with a persuasive argument for the use of omnibus tests, rather than of tests with factor scores, for the prediction of school achievement.

In all, the revised Henmon-Nelson impresses the reviewer as a scholarly example of the best in test construction, and as a remarkably efficient instrument for its length.—*L. F. S.*

Personal Adjustment: An Approach through the Study of Healthy Personality

By SIDNEY M. JOURARD, *Ph.D., Associate Professor of Psychology, University of Florida.* Written from a positive point of view, this unique work discusses the factors determining the healthy, adjusted personality. The author outlines each aspect of personality and compares the normal, unhealthy and healthy versions of the trait in question. The sections on expositions of interpersonal relations, love, the self-structure, and conscience are new to the field.

A pre-publication reader comments: "I think it is indeed a new approach to mental hygiene and a far better one than what we now have available." January, 1958, 462 pages, $5.50

INTRODUCTORY CLINICAL PSYCHOLOGY

By SOL L. GARFIELD, *University of Nebraska Medical School.* Stressing diagnosis, psychotherapy and research, this book presents a study of the development and current role of clinical psychology, both as a science and as a profession. *1957, 469 pages, $6.00*

DIFFERENTIAL PSYCHOLOGY: Individual and Group Differences in Behavior, Third Edition

By ANNE ANASTASI, *Fordham University.* Thoroughly revised to include recent American and European research in biology, anthropology and sociology, the third edition of this authoritative volume contains much significant new material. *1958, 664 pages, $7.50*

The Macmillan Company
60 FIFTH AVENUE, NEW YORK 11, N.Y.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

Further Evaluation of the Archimedes Spiral Aftereffect

Harold Gilberstadt, Jerome D. Schein
VA Hospital, Minneapolis, and University of Minnesota

and Albert Rosen
State College of Washington

The present investigation was stimulated by the report of Price and Deabler (5) who, after testing for the spiral aftereffect on groups of normal, nonorganic psychiatric, and organic brain-damaged hospitalized patients, concluded: "When failure to perceive the aftereffect is found it is almost certain that cortical involvement exists" (5, p. 301). They state that the results of a previous study by Freeman and Josey (1) led them to believe that this phenomenon might be used to develop a technique for the detection of organic brain damage. In a subsequent study, Gallese concluded that "the test almost always indicates organicity when organic scores are obtained although the converse is not true" (2, p. 257).

This study represents an effort to evaluate further the spiral aftereffect as a screening device for cortical brain damage with a VA GM&S hospital sample. In such a setting where large numbers of patients are examined for diagnosis and possible acute and intensive treatment, it would be highly desirable for the clinical psychologist to be able effectively to go beyond the mere assessment of the efficiency of functioning in those areas for which we presently have standardized techniques such as the Wechsler-Bellevue. There is little question that we are still far from able to state with any degree of confidence to what extent specific functional or organic variables contribute to the observed or predicted performance of the patients we examine. Yet it is apparent that our goal in the area of differential diagnosis must be to rely as little as possible on the art of clinical diagnosis which is strong in the tradition of neurology and psychiatry and to continually strive to develop rigorously standardized techniques from which statistically sound predictions of known validity can be generated.

The results obtained by Price and Deabler were of more than usual interest because, considering the complexity of the human organism and the nature of the variables involved in the loss of function from cortical brain damage, it is hard to overcome initially strong scepticism toward the claim that a single, brief test of one modality could develop into an almost infallible sign of such damage. As a consequence, the study by Price and Deabler seemed to demand further evaluation.

Method

Every effort was made to follow the instructions of Price and Deabler for administering the test. An Archimedes Spiral of 920°, or two and one-half turns about its center, was constructed by painting a black spiral on a white background. This wheel was affixed to a reversible motor with a variable speed control which was set at 100 rpm for the experiment. The apparatus was placed in an inside room which was well illuminated by fluorescent lighting. The spiral was placed at eye level and the subject was seated at a distance of eight feet from the apparatus.

One experimenter was employed to test all of the patients. Four trials were given alternating clockwise and counter-clockwise rotations. Scoring was on the basis of 1 or 0 for each trial. Samples of responses to the aftereffect scored 1 were: "It's coming toward me." "Getting larger." "Expanding." "Going in the opposite direction." "Swelling." "Shrinking."

"Opposite." Samples of 0 scoring were: "Nothing." "Stopped." "Standing there." "Standing still." "Same."

A check of scoring reliability by four judges on the first 35 cases showed no disagreements except for one instance in which a judge misread the handwriting on the recording form.

Sample

During a three-month experimental period all available patients admitted to the Psychiatry and Neurology Services respectively were tested. The total number of admissions to Psychiatry were 112 and to Neurology 175.

Table 1

Diagnoses and Scores on Archimedes Spiral Aftereffect of Psychiatric Sample

($N = 87$)

Diagnosis	Distribution of Scores				
	0	1	2	3	4
Organic ($N = 7$)					
Acute brain syndrome associated with alcohol intoxication	1				
Chronic brain syndrome with psychosis	1	1			
Encephalitis, epidemic chronic	1				
Encephalopathy, chronic progressive		1			
Depressive reaction with hypertension and cerebral vascular accident		1			
Alzheimer's disease	1				
Nonpsychotic ($N = 57$)					
Anxiety reaction			2	1	19
Conversion reaction					1
Depressive reaction	2	1	1	2	5
Dissociative reaction				1	2
Obsessive-compulsive reaction					4
Psychophysiologic reaction					4
Adult situational maladjustment					2
Chronic alcoholism					3
Dysocial reaction					1
Emotional instability reaction		1		1	1
Inadequate personality					1
Mental deficiency					1
Psychotic ($N = 23$)					
Involutional melancholia		1			1
Paranoid schizophrenia				3	9
Schizophrenic reaction other		1		2	5
Manic depressive					1

To maintain standard testing conditions, an initial requirement was instituted that only patients who were able to come to the testing room would be included; the criterion being the judgment of the nurse in charge of the ward. As a result of this condition, 16 patients were omitted from the Psychiatry sample. An inspection of the cases showed physical illness, assaultiveness, and suicidal risk to be the bases of the nurses' decisions. Seven patients were omitted through administrative error. Two patients who did come for testing were unable to attend to the task. Table 1 shows the distribution by diagnosis and score of the remaining 87 cases which constitute the Psychiatry sample.

Thirty-five patients admitted to Neurology were omitted from the sample of these cases. Fifteen were discharged before testing could be done, six expired, eleven were bedridden, three were aphasic and after reporting for the test did not have sufficient verbal ability to be tested under the standard conditions. Table 2 gives the distribution by diagnosis and score of the remaining 140 cases.

Mean age of the Psychiatric sample is 41.2 years and mean age of the Neurologic sample 43.4 years. Both groups were tested an average of five days after admission. Diagnoses in all cases were final diagnoses and were uncontaminated by test results which were withheld until all diagnoses had been obtained.

Results and Discussion

In reporting their findings, Price and Deabler used the terms "organic," "brain damaged," and "cortical" as though they were equivalent. That the first two terms may convey very similar meaning is understandable, but the term cortical has a highly specific connotation referring to a particular portion of the brain. Price and Deabler's summary statement led us to conclude that, despite the ambiguous use of these words, their intention was to restrict the test to the detection of *cortical* brain damage. It seemed imperative, then, to dichotomize our organic brain syndrome cases into cortical and noncortical subsamples. Parenthetically, it may be noted that assigning only cortical cases to the category of expected positives was more favorable to the hypotheses of Price and Deabler

Table 2

Diagnoses and Scores on Archimedes Spiral Aftereffect of Neurologic Sample

($N = 140$)

Diagnosis	Distribution of Scores				
	0	1	2	3	4
Organic brain syndrome[a] ($N = 68$)					
Traumatic brain injury	1		1	2	5
Demyelinating disease of brain				1	5
Brain tumor	2			1	1
Cerebral vascular accident	4			3	4
Convulsive disorder	2	1		2	15
Chronic brain syndrome	1		2		1
Meningitis					2
Paralysis agitans		1			3
Other					8
Organic non-brain syndrome[b] ($N = 56$)	9	3	2	8	34
Organic disease not found ($N = 8$)				1	7
Psychoneurosis[c] ($N = 8$)			2	1	5

[a] Specific diagnoses in each category are as follows: *traumatic brain injuries*—2 posttraumatic epilepsy; 1 cerebral concussion, 1 basilar skull fracture, 2 cerebral contusion, 1 penetration wound right parietal area, 1 postconcussion syndrome, 1 post-traumatic encephalopathy; *demyelinating diseases*—3 amyo-trophic lateral sclerosis, 1 multiple sclerosis, 1 primary lateral sclerosis, 1 pyramidal tract disease bilateral; *brain tumors*—1 brain tumor, 1 retrobulbar, 1 right temporal, 1 right thalamic; *cerebral vascular accidents*—1 c.v.a., 1 embolus middle cerebral artery with convulsive disorder, 1 occlusion of right internal carotid, 4 cerebral thrombosis, 1 internal carotid artery, 1 post cerebellar artery, 1 old with motor aphasia and right hemi-paresis, 1 right hemisphere; *convulsive disorders*—1 black-out spells, 1 cetrencephalic seizures, 6 grand mal epilepsy, 1 grand mal and petit mal epilepsy, 1 grand mal with drug withdrawal, 8 epilepsy of undetermined etiology, 2 psychomotor seizures; *chronic brain syndrome*—1 chronic alcoholism with secondary diffuse cortical atrophy, 1 degenerative brain disease, 2 unknown etiology diffuse; *meningitis*—1 pneumococcal type III, 1 syphilis tertiary; *other*—2 metastatic carcinoma, 1 cervical mass, 1 skull defect right frontal, 1 subarachnoid hemorrhage, 1 tabes dorsalis, 1 headache with retinal hemorrhages, 1 hypertensive cardiovascular disease.

[b] Specific diagnoses are as follows: 1 arachnoid cyst, 1 arteriosclerotic heart disease, 1 back-pain undetermined etiology, 2 Bell's palsy, 1 cervical spine trauma, 1 degenerative cord disease, 1 epididymal cyst, 1 ganglion cyst right wrist, 1 metastatic hypernephroma, 4 low back syndrome, 1 lumbar strain, 1 Meniere's disease, 1 myotonia dystrophica, 1 nephrotosis, 1 posttraumatic neuropathy, 24 herniated nucleus pulposus, 1 occlusion anterior spinal artery, 3 peripheral neuritis, 3 radiculitis, 1 rheumatic heart disease, 1 sacroiliac strain, 1 sciatic syndrome, 1 possible slipped disc, 2 syncope of undetermined etiology.

[c] Specific diagnoses are as follows: 1 anxiety reaction, 1 functional cephalgia, 4 conversion reaction, 1 depressive reaction, 1 psychoneurosis severe with paranoid trends.

than the inclusion of all brain-damage cases would have been.

We relied on the best judgments available to classify our sample on the basis of diagnosis.[1] Nevertheless, we did so with reservation, since a realistic consideration of the problem must lead to the conclusion that within the limits of a study of this type it is not possible to establish a very satisfactory criterion of cortical brain damage. Even with unlimited resources there still would be serious weaknesses in the criterion because of the difficulties inherent in the accurate determination of the localization and extent of cortical brain damage with even the best of current neurological techniques.

Price and Deabler did not elaborate on the basis by which they conclude that their cases of CNS syphilis, cerebral vascular accident, chronic brain syndrome associated with alcoholism, cerebral arteriosclerosis, posttraumatic encephalopathy, and advanced Parkinson's disease have known cortical involvement beyond stating that the diagnoses were based on hospital staff action. Because of our lack of confidence in such diagnoses and for the reason noted above, we were conservative and assumed that only the 30 cases, which included traumatic brain injury, brain tumor, cerebral vascular accident, chronic brain syndrome, and meningitis (listed under the organic brain syndrome subgroup in Table 2), plus the seven cases from the organic subgroup (listed in Table 1) have cortical brain damage of a pathological degree.

Meehl and Rosen's (3) procedure for the evaluation of any assessment or prediction device was followed in the analysis of the test results. In the first three columns of Table 3, f denotes the number of diagnosed cases of cortical damage receiving each score, cf indicates the cumulative frequency and P_1 represents the proportion of the total group of those with cortical damage who scored at or below each scoring interval. The next three columns give the same data for the group without cortical damage. In order to conserve space we have eliminated the intermediate calculations given for illustrative purposes by Meehl and Rosen. H_T represents the proportion of correct classifications for the given cutting line; i.e., the number of valid positives plus the number of valid negatives divided by the total number of cases.

When all cases are considered, it will be noted that the test does no better than the base rates; i.e., if we simply called all cases

[1] We wish to thank E. T. Bell, Emeritus Professor of Pathology, University of Minnesota Medical School, and David Webster of Minneapolis VA Hospital neurology staff for their assistance in categorizing the neurological cases.

Table 3

Cumulative Frequency Tables for Calculation of Optimal Cutting Scores for Archimedes Spiral Aftereffect

		Cortical			Noncortical			
	Score	*f*	*cf*	P_1	*f*	*cf*	P_2	H_T
Combined sample[a]	0	12	12	.3243	13	13	.0684	.8344
	1	3	15	.4054	9	22	.1158	.8076
	2	3	18	.4865	7	29	.1526	.7896
	3	6	24	.6486	23	52	.2737	.7139
	4	13	37	1.0000	138	190	1.0000	.1600
Psychiatry cases[b]	0	4	4	.5714	2	2	.0250	.9427
	1	3	7	1.0000	4	6	.0750	.9310
	2				3	9	.1125	.8965
	3				10	19	.2375	.7815
	4				61	80	1.0000	.0800
Neurology cases[c]	0	8	8	.2667	11	11	.1000	.7670
	1	0	8	.2667	5	16	.1455	.7311
	2	3	11	.3667	4	20	.1818	.7234
	3	6	17	.5667	13	33	.3000	.6720
	4	13	30	1.0000	77	110	1.0000	.2100

[a] Base rate for cortical brain damage in this sample is .16.
[b] Base rate for cortical brain damage in this sample is .08.
[c] Base rate for cortical brain damage in this sample is .21.

free of cortical damage we would correctly identify 84% of the sample. Using the test, we correctly identify 83.4%, provided that we use a cutting score between 0 and 1. Using the cutting line between 2 and 3 suggested by Price and Deabler, we would classify the cases correctly only 78.9% of the time as compared to the base rates of 84%.

In Psychiatry, where 8% of the sample were diagnosed as having cortical damage, if the cutting line were set between 1 and 2, 93% of the cases would be correctly identified as compared to 92%, using only the base rate and calling all cases non-brain-damaged. Maximum effectiveness is again attained with a cutting line between 0 and 1 where 94.3% of the cases are correctly identified.

In the Neurologic sample, the test is unable to improve upon the base rates. By routinely calling all cases not cortically damaged, 79% would be correctly classified. Using a cutting score between 0 and 1 the test would attain its maximum number of correct identifications, 77%. Use of the test, then, would result in a 2% *decrement* in predictive accuracy.

In addition to the total number of correct classifications, evaluation of a clinical instrument is dependent upon the importance attributed to detecting the greatest number of cortically damaged cases and to the cost and the facilities available for further neurological evaluation of mislabeled, noncortically damaged patients. If one wished to avoid missing any cortical cases regardless of the cost, he would choose a cutting line between 3 and 4 for a population resembling our combined Neurology and Psychiatry samples. This would yield the detection of 64.9% (P_1) of the cortical cases at the expense of misclassifying 27% (P_2) of the noncortical cases. If one were interested in avoiding additional neurological examinations because of staff limitations, he would select a cutting line between 0 and 1 where 32.4% of the cortical cases are identified, and only 6.8% of the noncortical cases are misdiagnosed by the test.

Table 4 contains a similar analysis of the data presented by Price and Deabler (5), Gallese (2), and Page, Rakita, Kaplan, and Smith (4). We have also repeated the data for our combined Neurology and Psychiatry samples to simplify comparison of all four studies. Since none of the other three studies report the base rates for their populations, we have used the empirical rate obtained from our combined sample (.16) as the estimate with which to make the necessary calculations.

It will be noted in Table 4 that the maximum total hit rate (H_T) varies from 96.4% in the Price and Deabler study, to 90.4% for Gallese, to 81.0% for Page *et al.*, to 83.4% for the Minneapolis sample, and that the cutting line at which the maximum is obtained alternates from a point between 2 and 3 in the first two studies to between 0 and 1 for the last two studies. Gallese stated that his method was more apt to reduce the proportion of false positives inasmuch as he continued to press for a "normal" response if it were not at first obtained. He was also more liberal in scoring responses in the "normal" direction. In the Page *et al.* study, the spiral was revolved in one direction only for three trials rather than in two directions for four trials.

The present study, however, does not differ from Price and Deabler's in test administration. The reduction in total hit rate to 83.4% (using a cutting line between 0 and 1) is in our opinion largely a result of differences in sampling. Our sample was drawn randomly. All consecutive admissions who could be tested were tested, with exceptions as noted above. Furthermore, our patients were probably less deteriorated than those selected by Price and Deabler though we cannot be certain because of the paucity of information they provided. We feel that the lack of description of their sample represents a very serious deficiency in the usefulness of their results. It is possible to select samples of brain-damaged and psychotic populations such that the brain-damaged, being severely deteriorated, can fail almost any task given them and the psychotics, being relatively intact, obtain a high rate of success. Such a procedure yields a "mythical population," i.e., one which does not occur in practice. Tests validated on such groups will give deceptively good discrimination between groups, only to fail on cross-validation.

Table 4

Percentage of Hits for Each Cutting Line in Different Hospital Samples Assuming a Base Rate of .16

	Percentage of Hits in Each Study			
Cutting Lines Between	Price & Deabler	Gallese	Page *et al.*	Minneapolis
0 and 1				
H_P (Positive hit rate)	100.0	*	43.2	47.4
H_N (Negative hit rate)	92.9	*	91.8	87.9
H_T (Total hit rate)	93.6	*	81.0	83.4
1 and 2				
H_P	91.5	*	*	40.0
H_N	94.7	*	*	88.6
H_T	94.3	*	*	80.8
2 and 3				
H_P	87.4	72.0	*	37.8
H_N	98.2	93.6	*	89.6
H_T	96.4	90.4	*	79.0
3 and 4				
H_P	78.9	*	*	31.1
H_N	99.7	*	*	91.6
H_T	95.5	*	*	71.4

* Data unavailable.

It is perhaps well to note that developing a mythical population is not entirely without virtue. If the means of obtaining such a sample are carefully reported, it may be possible to locate or duplicate a reasonably similar sample. On the other hand, nonrandomly selecting a sample may result in one being unable to validly predict what the results might be even in one's own institutional setting where the sample was drawn. We can only speculate, therefore, as to the nature of the differences between the other samples and the one drawn for this study. Gallese, for example, reports that the median duration of hospitalization for his organic group is 65 months and for his schizophrenic group 33 months in contrast to our groups who were tested an average of five days after admission to the hospital. There are also substantial differences in the percentage of cases of CNS syphilis in our samples and those of Gallese, and it is probable that this would be the case in a comparison with the Price and Deabler sample. The fact that Gallese obtains results intermediate between ours and those of Price and Deabler might suggest that he has drawn a sample which more closely approximates theirs on such significant variables.

Table 4 also contains for each cutting line the positive hit rate

$$\left(H_P = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}\right)$$

and the negative hit rate

$$\left(H_N = \frac{\text{true negatives}}{\text{true negatives} + \text{false negatives}}\right).$$

Thus, in Price and Deabler's study, 100% of cases obtaining a score of 0 are diagnosed as cortically damaged, while this is true for only 43.2% of the Page *et al.* sample and 47.4% of our sample. Conversely, if we wished to maximize the negative hit rate, the cutting line should be placed between 3 and 4 where the rate of correct negative predictions was 99.7% for the Price and Deabler sample and 91.6% in our sample. One can, then, select a cutting line depending on (*a*) which sample most closely resembles the one in which the test is to be used and (*b*) whether one wishes to maximize H_P, H_N, or H_T.

In most clinics, use of the spiral aftereffect as a screening device will be considered in relation to the efficiency of the instrument as well as the kinds of decisions which will be made on the basis of positive results. Assuming that further cross-validation will not cause much shrinkage, it might be practical to use the test with samples which resemble our Psychiatric sample where only 8% false positives were obtained with a cutting score between 0 and 1. Considering the small amount of time and effort involved, it would be feasible to administer the test to all patients admitted to Psychiatry and to recommend a neurological workup for those who score below this cutting line.

The test would not seem to have utility for the screening of Neurology patients. Since there is such an unfavorable ratio between valid and false positives it would not appear feasible even to advise the neurologist to use the test findings as an added weight in his diagnostic formulation.

Summary

Evaluation of 87 consecutive testable admissions to the Psychiatry Service and 140 to the Neurology Service of the Minneapolis Veterans Administration Hospital by means of the Archimedes Spiral Aftereffect indicated that the test had very limited clinical usefulness for detecting pathological cortical brain damage. Use of the test did not significantly improve the diagnostic efficiency which would be obtained by using the base rates.

Problems of sampling and criteria were discussed. Differences between this study and those previously reported appear to be largely a function of differences in the samples studied. Problems inherent in the use of nonrandom samples for purposes of validation were also pointed out.

Received August 21, 1957.

References

1. Freeman, E., & Josey, W. E. Quantitative visual index to memory impairment. *Arch. Neurol. Psychiat.*, 1949, **62**, 794–796.
2. Gallese, A. J. Spiral aftereffect as a test of organic brain damage. *J. clin. Psychol.*, 1956, **12**, 254–258.
3. Meehl, P. E., & Rosen, A. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychol. Bull.*, 1955, **52**, 194–216.
4. Page, H. A., Rakita, G., Kaplan, H. K., & Smith, N. B. Another application of the spiral aftereffect in the determination of brain damage. *J. consult. Psychol.*, 1957, **21**, 89–91.
5. Price, A. C., & Deabler, H. L. Diagnosis of organicity by means of spiral aftereffect. *J. consult. Psychol.*, 1955, **19**, 299–302.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

The Spiral Aftereffect in a Neurological Setting

David Berger,[1] Richard Everson, Louis Rutledge[2]
VA Hospital, Leech Farm Road, Pittsburgh

and Yale David Koskoff[3]
Montefiore Hospital, Pittsburgh

The spiral aftereffect may normally be seen when a circular disk supporting an Archimedes spiral is rotated within certain limits of speed. During the period of rotation, the spiral appears either to expand or to contract, depending on both the direction of rotation and the direction of the coil of the spiral. When the rotating movement of the disk is interrupted, an aftereffect occurs so that the spiral appears to rotate in the opposite direction.

Like many techniques being reactivated in current perception research, the spiral aftereffect has a substantial history. In 1850, Plateau, the Belgian scientist, reported the first systematic study of the moving spiral. More lately, the relation between brightness, direction, velocity, etc., and the perception of the aftereffect has been investigated (3, 9).

Along with the interest in the various parameters of the experience is the attempt to relate the performance on the spiral to discrete mental functions and to brain damage. Freeman and Josey's paper (1) on the correspondence between memory impairment and reduced ability to perceive the aftereffect, spurred Price and Deabler (7) to measure the relationship between the aftereffect and cortical damage. The results of this latter effort are most impressive in that virtually no overlap in performance was found between a group of brain-damaged patients and two control groups. In the brain-damaged group 98% failed to perceive the aftereffect in each of the four consecutive trials, while 95% of the nonorganic psychiatric population, and 92.5% of the normal population, were able to perform satisfactorily on each of the four trials. In a subsequent paper, Garrett, Price, and Deabler (2) repeated their original work with 40 chronic brain-injured patients. Their findings substantiate their previous results, inasmuch as 97.5% of the sample failed to perceive the aftereffect in each of the four trials. Page *et al.* (6) have provided another test relating performance on the spiral to cortical damage. While their results were in the same direction as Price and Deabler's, the ability of the spiral test to discriminate was patently less efficient. In another study conducted by Smith and Goldberg,[4] the spiral was administered to normals, psychiatric patients, organics, and postshock patients. The test proved capable of discriminating between pathological processes and normals, but apparently failed to distinguish organic impairment from the functional disorders.

The present study was undertaken to cross-validate the findings reported by Price and Deabler and particularly the statement, "While adequate explanation of the phenomenon remains to be hypothesized and verified, the results are such as to justify present clinical use of this technique in diagnosis of organicity (7, p. 302). In addition, the following problems were evaluated: (*a*) the effect of the direction of the coil of the spiral on difficulty and duration of the after-

---

[1] Now at Crile VA Hospital, Medical Service.

[2] Now at VA Hospital, Denver.

[3] The authors wish to express their appreciation to Joen Fagan, graduate student at Pennsylvania State University, and Ted Rice, graduate student at Western Reserve University, who assisted in collecting and evaluating the data.

[4] P. Smith and L. Goldberg. Personal communication. April, 1957.

Bureau Educ. Psy. Research

effect, (*b*) the relation between visual acuity and the aftereffect, and (*c*) the relation between the aftereffect and various neurological indicia.

Method

Subjects

With the exception of bedfast patients and those patients who were too incapacitated to cooperate, all patients hospitalized during a three-month period on a neurological ward in a VA hospital were selected for study. Some months after the experimental data were collected, the patients' clinical charts were reviewed for information relevant to the following areas: diagnoses,[5] admission symptoms, electroencephalographic (EEG) reports, pneumoencephalographic studies, skull X-ray reports, findings on spinal fluid examination, and reports on visual fields. In certain cases, the chart failed to yield data for each variable. The availability of this type of information was largely a function of each patient's condition and the consequent need for laboratory study.

Apparatus

A six-inch Archimedes spiral formed in 2½ convolutions was drawn and photographed. Two identical spirals were developed, except that the second was the photographic reversal of the first. The two spirals were then mounted on two 6-in., circular, metal disks. A small metal collar was centered and fastened to the reverse side of the disks. The collar was threaded to complement a machine screw which had been fitted to the center post of a phonograph turntable. The turntable, together with its driving mechanism, was supported on an adjustable laboratory stand so that the target spiral could be presented at eye level to a subject (*S*) seated at a distance of 8 ft. The turntable was activated by a 78 rpm unidirectional motor. Spiral A seemed to expand while rotating and normally would be said to contract in the motionless aftereffect condition. Spiral B, since it was the photographic reversal, and since the motor was unidirectional, appeared to be contracting while rotating and to be expanding in the motionless aftereffect condition.

In addition to the spiral apparatus, a Snellen chart was used to assess visual acuity. As with most work on visual acuity, the minimal separable or visual angle was considered the invariant. However, since the influence of distance is not clearly defined in present-day knowledge of visual acuity, it was also controlled. The size of the room used for testing was such that the *S*s had to read the chart from a distance of 18 ft. By convention, readings of visual acuity, thus obtained, would be adjusted to comply with a 20-ft. distance from the chart. Since the distance error in the ratio was constant in this study, it was not necessary to make this adjustment. In all calculations based on Snellen chart readings, the denominator in the ratio was used exclusively (i.e., with readings such as 18/30 or 18/20, the computations were performed only on the denominators).

Testing Procedure

Two experimenters (*E*s) were present during all testing. One *E* typically discussed the procedure with *S* and recorded behavior. The second *E* observed and adjusted the apparatus. The *S*s were encouraged to wear glasses. When glasses were used, they were required throughout the testing. All visual tasks in this experimentation were viewed binocularly under artificial overhead light.

The patient was told that the tests about to be administered were a special eye examination. The Snellen chart was then called to the patient's attention. He was told to read each succeeding line from left to right until interrupted. The score on this task was the denominator of the acuity ratio of the most difficult line read faultlessly. The criterion required that the line be read perfectly twice and that the succeeding line be faulted twice. The latter condition was obviously inapplicable in the case of an errorless performance on the entire chart. The *S*s were prompted to rest their eyes whenever a line was failed.

The spiral task was then introduced. The two spirals were presented twice each for every *S* in "controlled" random order for a total of four exposures per *S*. The order of presentation was thus randonized for blocks of six patients: ABAB; ABBA; AABB; BBAA; BABA; BAAB. Hereafter, each of the four exposures are referred to as a presentation. Between presentations, a cardboard shield was used to cover the apparatus and the *S* was instructed to close his eyes to rest them. Every effort was made to make the interval between presentations uniform both for time and auditory cues. This was necessary in order to reduce the extraneous information which might be used to anticipate which spiral was being readied for the next presentation.

---

[5] All diagnoses reported in this study were established after weeks of observation and study.

The instructions for the spiral task were, "Now please look at the center of the design on the wheel and keep looking right at the center." At that point, the motor which activated the spiral was switched on. After 10 seconds, the *S* was asked to describe what he saw. All *S*s included in the experimental sample were able to describe some form of expanding or contracting experience while the spiral was rotating. The form described depended, of course, on which of the two spirals was being used during the initial presentation. After thirty seconds of observation, the power was cut and the spiral came to rest. The average temporal lag between the time the power was cut and the time the spiral came to a complete stop was 2½ seconds. During this interval, *S* was cautioned again to "Keep your eyes on the center of the design on the wheel." After five seconds elapsed the *S* was asked a series of questions in the following sequence: (*a*) What do you see now? (*b*) Does it look different now? (*c*) Does it look like it's moving now? (*d*) Does it look like it's getting larger or smaller now?

The procedure was in marked contrast to the one employed by Price and Deabler. They assert "there was little ambiguity as to the presence or absence of the aftereffect with most *S*s" (7, p. 302). In the pretesting of the apparatus used in the present study, there were many occasions when *S*'s verbal report failed to establish the legitimacy of the aftereffect experience. The graduated series of questions was designed to uncover the essence of *S*'s perceptual experience. A response at any level of questioning which indicated that the *S* experienced a change either in direction, size, or in distance was acceptable as long as the experience was in the expected direction (i.e., with a spiral which would normally be seen as contracting in the aftereffect condition, reports such as moving away, increasing distance, or decreasing size were all equally "correct"). All acceptable responses were scored plus. When *S*'s response for any presentation was "correct" at any one of the four levels of questioning, that presentation was completed. Thus, all four levels of questioning were used only for responses which were not sufficiently clear to be scored with confidence. A response which ultimately failed to convince the *E* that the *S* perceived the aftereffect or which, at any level of questioning, described the movement in the "wrong" direction, was scored as a minus performance for that presentation. Allowedly, the verbiage employed in the questions tended to promote ideation about movement, distance, and size. Nevertheless, a "correct" response had to incorporate direction of change. In spite of the nature of the questions, the direction of change had a chance expectancy of two correct guesses in four presentations. Accordingly, the *S* was classified as a "Seer" if he scored plus in at least three of the four presentations. *S*s earning two or more minus scores were classified as "Nonseers."

In the case of "Nonseers," the fourth presentation completed their performance. The "Seers," however, were asked to continue. The particular spiral used in the fourth presentation was retained through four additional trials, during which the duration of the aftereffect was ascertained. These four duration exposures will hereafter be referred to as trials. The time per trial was recorded in seconds from the instant the spiral was set into motion, through the 30-second exposure period, and up to the moment the *S* indicated that the aftereffect seemed to be extinguished. The directions for these four trials were, "Do you remember just a moment ago when we stopped the wheel, you told me that the design on the wheel looked like it was '____________' (the subject's own descriptive phrase was used). Now when we start the wheel this time, look at the center again and keep looking at it after I stop it. Then tell me when the '____________' (inserting the same phrase again) first stops." After completing four trials, the "Seers" were excused from the test situation.

Results

The final sample tested numbered 110 patients, all of whom were admitted to the hospital ostensibly for the diagnosis and treatment of neurological conditions. The mean age of the sample was 38.4 years, with an *SD* of 10.99 years and a range of 21 to 64. The primary diagnoses represented in the sample are presented in Table 1 in order to provide some notion of the character of the ward population at the time the testing was conducted. Of the 110 patients tested, 31 (28%) were classified by their performance as "Nonseers." The incidence of "Nonseers" for the various diagnoses is also presented in Table 1.

Comparisons between "Seers" and "Nonseers"

The means of the ages for the "Seers" and "Nonseers" were contrasted, and the lack of significance is summarized in Table 2. In order to compare the means of the Snellen chart performances of the "Seers" and "Nonseers," a square root transformation was used since the distributions were rather severely skewed. The results of this comparison as presented in Table 2 were significant ($< .01$). From this finding it may be said that "Nonseers" have less acute vision than "Seers" and that this relationship does not appear to be related to age.

Comparison of the Two Spirals

The distributions of duration times for each of the two spirals also required normalizing.

Table 1

Diagnoses and Incidence of "Nonseers"

Diagnosis	*N*	"Nonseers"	Diagnosis	*N*	"Nonseers"
Multiple sclerosis	24	14	Syncope	3	0
Encephalopathy	15	2	Arteriosclerosis	3	1
Epilepsy, grand mal	15	2	Intracranial aneurysm	3	1
Parkinsonism	7	4	Cerebral thrombosis	3	0
Neurotic reaction	7	0	Alcoholism, chronic	2	0
Epilepsy,			Cephalgia	2	0
Jacksonian	4	3	Amyotrophic lateral sclerosis	2	0
psychomotor	4	0			
idiopathic	3	0			

Note.—There was also one each of the following diagnoses: Muscular dystrophy, migraine, toxic neuropathy,* papillitus optic nerve, cerebral embolism,* Friedrich's ataxia, poliomyelitis, progressive spinal muscular atrophy, neuralgia, Huntington's chorea,* osteoarthritis, transverse myelitis, and schizophrenic reaction, paranoid type.*

* Denotes a "Nonseer."

Again the square root transformation was employed. The results of the comparison of the mean duration times obtained on the two spirals are presented in Table 3. It is to be remembered that the duration trials were administered only to "Seers." Accordingly, Spiral B, which normally is seen as expanding in the motionless condition, is seen for a significantly longer time than is Spiral A ($< .001$).

Next, the relative difficulty of the two spirals was examined. Spiral A (contracting aftereffect) yielded 60 failures (27%) out of 220 exposures. Spiral B yielded 36 failures (16%) in a similar number of exposures. These percentages were tested by the critical ratio formula and proved quite significant ($< .01$). Thus Spiral B (expanding aftereffect) was appreciably less difficult than Spiral A. The performance on the two spirals was then distinguishable both for duration and difficulty of the aftereffect. The aftereffect on Spiral B was easier to see and was seen for a longer time interval.

Relation Between Spiral and Clinical Criteria

The information gathered from the patients' clinical charts was related to the performance on the spiral aftereffect. This approach involved analyzing the chart data on five independent measures, as well as via a global impression of brain damage. The five indices used were EEGs, pneumoencephalograms, skull X rays, spinal fluid, and visual fields. The global judgments integrated these five indices, as well as the admission symptoms and the diagnoses into a single, aggregate judgment.

Two judges independently evaluated the abstracted chart summaries for each of the five indices with 20 of the patients. These 20 charts were selected in a random manner with the one stipulation that information on the five indices was available. The judges were in-

Table 2

Age and Visual Acuity in "Seers" and "Nonseers"

	Mean scores		*SD*			
Variable	"Seers"	"Nonseers"	"Seers"	"Nonseers"	*t*	*p*
Age	37.71	40.26	10.41	12.35	1.00	—
Visual acuity[a]	4.4871	5.9007	1.3021	2.4707	2.98	<.01

[a] Normalized via a square root transformation.

Table 3

Comparison of Duration Time[a] on Spirals A and B

Spiral	N	Mean	SD	t	p
A	152	6.3861	.6001	6.2969	$<.001$
B	163[b]	6.8879	.8018		

[a] Normalized via a square root transformation.
[b] One patient failed to perceive the aftereffect on one of the duration trials on Spiral B.

structed to assign each case to one of the three categories: obvious brain damage, questionable brain damage, or no evidence of brain damage for each variable (EEG, etc.). The two judges agreed perfectly on 84% of the ratings. With this degree of consistency, it was possible to use the judgments of a single judge for the purpose of relating the five indices to the spiral performance. Whenever data were not available, the patient was excluded from the analysis on that particular index. The contingency coefficients between the spiral performance and the neurological measures, as well as the levels of significance, are presented in Table 4. Only chance association was demonstrated between the performance on the spiral test and the three clinical variables: EEGs, pneumoencephalograms, and skull X rays. However, in the case of the visual field examinations and the spiral fluid, significant correlations (both $< .05$) were obtained with the performance in the aftereffect situation. The global estimate of cortical damage was made by one of the authors.[6] The contingency coefficients along with the level of significance ($< .05$) are also presented in Table 4.

Discussion

When certain stimulus patterns are observed in motion at critical speeds, the activity may appear to continue for some appreciable time after the objective motion ceases. Such experience is often termed motion afterimage. The spiral aftereffect is a form of subjective motion afterimage, complicated by a negative experience of changing size or depth. As such, it incorporates radial movement, apparent movement toward or away from an observer, and gamma movement, another form of apparent movement usually associated with a swelling of the size of a suddenly illuminated stimulus.

The Spiral Aftereffect as a Clinical Tool

The results of the present study tend to confirm the general usefulness of the spiral aftereffect as a heuristic laboratory technique for studying brain pathology. There is, however, some real question about the propriety of the technique in clinical practice. The spiral task as explored in the present study does not appear to live up to the claims made for it by Price and Deabler as an individual diagnostic measure.

The critical need in clinical practice is for a technique sufficiently valid to be utilized with ambiguous and difficult problems in the diagnosis of brain damage. When performance on the spiral was compared with patients' diagnosis in the present study, the efficacy of the aftereffect as an individual diagnostic instrument seemed dubious. The disparities between the results in the studies by Price and Deabler and the present work may be attributable to many factors. For one thing, the degree of obviousness of brain dysfunction in the organic samples tested may have been at variance. In addition, some very real differences in methodology were employed. The scoring scheme in the present study made no distinction between verbal reports of changes in size and depth. In the study by Price and Deabler, size changes were given greater recognition in scoring. The justification for each specialization in scoring was not apparent to the present authors and, consequently, any response which indicated

Table 4

Relationship Between Performance on the Spiral Task and Clinical Neurological Tests

Variable	Contingency coefficient	Significance
EEG	.11	—
Pneumoencephalogram	.03	—
Skull X ray	.05	—
Spinal fluid	.24	.05
Visual fields	.27	.05
Global analysis	.25	.05

---

[6] These judgments were made by the neurologist cooperating on this project.

an appreciation of negative perceptual experience was fully accredited. In addition, the intention in the present study was to provide *S* with every chance to report an experience which may normally appear somewhat freakish and unnatural. Such an experience, it was reasoned, might be rather difficult to verbalize. Frequently it was found that *S*s did quite well on the task only after they fully understood the inquiry. At times this required the complete battery of four questions in order to elicit a satisfactory report. It is suggested that some of the inability of the organically impaired to report the aftereffect in Price and Deabler's paper may have been in part a consequence of communication failure rather than perception failure.

The visual acuity findings are further reason for doubting the legitimacy of the spiral aftereffect as a clinical measure of brain damage. The need to consider visual acuity in a study of perceptual processes is discussed by Yates (10) who admonishes against the omission of a measure of visual acuity in the development of perceptual tests of brain damage. This is an especially pertinent recommendation in instances where the experimental sample includes disease processes such as multiple sclerosis, wherein repeated atrophy of the optic nerve may frequently occur (4, 8). When atrophy develops, it seems to be present most often in the area of the maculopapillar bundle, which is involved in central vision. Since the spiral task involves foveal vision, in a large measure, the need for an index of visual acuity is apparent.

In a superficial attempt to explore further the relationship between visual acuity and brain damage, the Snellen chart readings were contrasted in the three groups differentiated in the global neurological appraisal. No relationship was found between these estimates of brain damage and visual acuity. The question as to the nature of the relationship between visual acuity and performance on the spiral must come under closer scrutiny before the technique can be used with confidence in clinical practice.

The Comparison of the Two Spirals

The performance on the two spirals indicates that the level of difficulty in perceiving the aftereffect, as well as the length of the aftereffect, was quite distinct. One of the best documented facts about the perseveration of movement afterimage is its dependence upon the velocity of the revolving stimulus. However, this factor was controlled in the present study, except for changes in line voltage which should have affected both spirals without bias. The finding that a spiral which seems to be expanding has a longer period of after-activity is consistent with observations of duration in gamma movement studies. In the latter studies, the expansion phenomenon, which occurs when a light source is introduced in a visual field, has been contrasted with the shrinkage which is reported when the light is dimmed or extinguished. As in the observations on the two spirals, the spreading movement occupies a significantly longer time interval than the shrinking movement. The two spirals were also compared to determine if they were equally effective in differentiating cortical damage as determined by the global clinical judgments. In spite of the fact that Spiral A elicited a greater absolute number of failures, there was no difference in efficiency of the two spirals in detecting brain damage.

The Spiral Aftereffect and Vision

The degree of relationship between the spiral performance and clinical visual tests inspired conjecture about the possible connection between deficit on the spiral task and injury to areas of the optic apparatus. The fact that the task was mediated through visual structures added to the temptation to localize the brain damage. Accordingly, all the patients whose clinical charts suggested a global evaluation of obvious or questionable brain damage were re-evaluated in an effort to estimate the localization of the presumed damage. However, the level of reliability for these judgments proved unsatisfactory, and the effort was abandoned for the present paper.

The evidence in this study suggested rather pointedly that two somewhat antagonistic aspects of visual functioning were simultaneously impaired in "Nonseers." On the one hand the ability of the "Nonseers" to make fine visual discriminations (visual acuity) was notably lessened. Such a function, according to various theories of vision, should require a high degree of insulation among juxtaposed individual pathways as well as a great number and superior differentiation of retinal receptors. The aftereffect task, on the other hand, would require either centrally or in the peripheral system rather complex neural interaction. Thus the dual failure of "Non-seers" on visual acuity and spiral aftereffect tasks provides concurrent behavioral evidence of deficit in anatomical insulation and physiological interaction.

Conclusions

The purpose of this study was to evaluate the use of the spiral aftereffect as an index of brain damage in a neurological setting. In addition, certain neurological specifications of organic impairment were employed as criteria against which perceptual experience was measured. And, finally, the direction of the coil of the spiral was varied, and the visual acuity of the subjects was tested in order to evaluate their effect upon performance. The results are as follows:

1. Of a sample of 110 patients admitted to a neurological ward, 28% were unable to perform satisfactorily on the spiral task.
2. Those who performed satisfactorily on the spiral task had significantly better visual acuity as measured by Snellen chart performance, but were not distinguishable in age from those whose performance on the spiral task was considered unsatisfactory.
3. Whether the aftereffect took the form of expansion or contraction was important for the ease with which it was reported, as well as for the length of time it was subjectively perceived.
4. Performance on the spiral was related to spinal fluid examinations, visual field studies, and global judgments of the likelihood of brain damage. Performance on the spiral was not related to EEG reports, pneumoencephalographic studies, or skull X-ray reports.

The degree of relationship demonstrated between estimated brain damage and spiral performance was taken as vindicating further exploration of the technique. It was suggested, however, that caution be exercised in the clinical use of the test, particularly in those situations in which the demand for such a test would be the most imperative, i.e., in the evaluation of individual cases where the diagnosis is less than straight-forward and direct.

Received August 5, 1957.

References

1. Freeman, E., & Josey, W. E. Quantitative visual index to memory impairment. *Arch. Neurol. Psychiat.*, 1949, **62**, 794–796.
2. Garrett, E. S., Price, A. C., & Deabler, H. L. Diagnostic testing for cortical brain impairment. *Arch. Neurol. Psychiat.*, 1957, **77**, 223–225.
3. Gates, L. W. The aftereffect of visually observed movement. *Amer. J. Psychol.*, 1934, **46**, 34–46.
4. Grinker, R. R., & Bucy, P. C. *Neurology*. Springfield, Ill.: Charles C Thomas, 1949.
5. Kohler, W., & Wallach, H. Figural aftereffect: An investigation of visual processes. *Proc. Amer. Phil. Soc.*, 1944, **88**, 269–357.
6. Page, H. A., Rakita, G., Kaplan, H. K., & Smith, N. B. Another application of the spiral aftereffect in the determination of brain damage. *J. consult. Psychol.*, 1957, **21**, 89–91.
7. Price, A. C., & Deabler, H. L. Diagnosis of organicity by means of the spiral aftereffect. *J. consult. Psychol.*, 1955, **19**, 299–302.
8. Schumacher, G. A. Multiple sclerosis and its treatment. *J. Amer. med. Ass.*, 1950, **143**, 1059–1065.
9. Wohlegemuth, A. On the aftereffect of seen movement. *Brit. J. Psychol., Monogr. Suppl.*, 1911, **1**, 1–117.
10. Yates, A. J. The validity of some psychological tests of brain damage. *Psychol. Bull.*, 1954, **41**, 359–379.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

A Further Study of the Perception of Apparent Motion by Schizophrenics

Rayford T. Saucer
VA Hospital, Perry Point, Maryland

The perception of visual apparent movement in normal subjects has been widely studied. It is known that when two lights are exposed alternately at a low rate of alternation, the perceptual emergent will be that of two lights flickering alternately. If the rate of alternation is raised, the subject will report a single light which swings smoothly back and forth. This state is known as optimal movement or beta movement (5). If the frequency of alternation is again raised, the subject will report seeing two lights which flicker simultaneously. It has also been noted that at the boundary between the disappearance of beta motion and the perception of simultaneous flicker, a formless dark object in motion may be perceived, called omega movement (4).

Research (1, 2, 3, 7) indicated that one of the most common results of organic damage to the cortex is the loss or impairment of the ability to perceive visual apparent movement. Saucer and Deabler (5), in an experiment designed to explore the use of beta movement as a diagnostic procedure in this area, used the appearance of omega movement as an index of failure to perceive optimal movement. They found that normal or control subjects were able to perceive beta movement at a much higher rate than organics.

They also found that a group of undifferentiated schizophrenics showed some degree of impairment in the ability to perceive optimal movement, although a group of lobotomized schizophrenics showed no measurable deficit. The present paper is a report of an experiment designed to further explore the loss of ability to perceive optimal movement in schizophrenic individuals.

The apparatus used in this study was a low frequency, square-wave generator, with a Hayden motor sweeping from 2.5 to 10 cycles per second in three minutes. The square-wave generator gated two Sylvania R1130B glow modulator tubes 180° out of phase. A green "Dialite" fresnel lens (No. 34–841) was placed over the end of the R1130B tube. The fresnel lenses were 2.5 inches center to center and 8 inches behind a $\frac{1}{8}$-inch opal Lucite screen 12 × 12 inches square. The intensity of the lights was adjusted to optimal movement for all subjects at 2.5 cycles per second and a mean breakdown rate of beta movement for a pilot group of controls of about 5 cycles per second.

Subjects and Method

Subjects (*S*s) were 21 undifferentiated schizophrenics, 21 paranoid schizophrenics, and 10 control *S*s. The patients were drawn from the NP population of Perry Point Veterans Administration Hospital, while the control group was composed of hospital nursing assistants and other nonprofessional hospital personnel. Patients were selected by ward psychiatrists and their diagnosis made known after they were tested.

The *S*s were brought one at a time into the testing room and asked if they would consent to sitting in a dark room for a few minutes. Those who were mute or incooperative were released immediately. The remaining *S*s were given three minutes of dark adaptation, during which time the experimenter engaged them in trivial conversation.

At the end of the three minutes, the stimulus lights were turned on at an alternation frequency of 2.5 cycles per second and the *S* asked to tell the experimenter what he saw. Almost invariably the *S* reported seeing a single light moving back and forth across the screen. The sweep drive was then turned on and the *S* asked to report any changes in the light. Most *S*s immediately reported "it's getting faster," but a simple "yes" seemed to be an adequate explanation.

When the *S* reported a dark object in motion, the sweep drive was stopped and the frequency in cycles per second read from a calibrated dial. If the *S* did not report a single light moving back and forth which was replaced by a dark object in motion, no score was given.

Results

The rate at which omega movement appeared was taken as the rate at which optimal movement had ceased. The mean rates and their associated standard deviations are given in Table 1.

The results are presented in Table 2 in terms of *t* and the Mann-Whitney *U* statistic (6, pp. 116–127). Although the differences between the schizophrenic group and the paranoid and control groups are statistically significant on the basis of the computed *t*, nonparametric analysis was also carried out because (*a*) it is evident that the variance is correlated with the mean, and (*b*) nothing is known concerning the population distribution of omega thresholds. The Mann-Whitney x/sigma values are equal to critical ratios when *N* is sufficiently large and hence may be interpreted by reference to a table of x/sigma for a normal distribution. It can be seen that the values are sufficiently large to allow the null hypothesis of no true difference between the schizophrenic group and the control and paranoid groups to be rejected at a comfortable level of significance.

Table 1

Threshold Means and Standard Deviations of Groups

Group	Mean	*SD*
Schizophrenic	3.428	.766
Control	5.15	1.39
Paranoid	5.417	1.96

Since the statistical analysis does not clearly show the degree of overlapping between the different categories, an analysis by approximate quartiles is given in Table 3. It can be seen from this table that whereas the majority of the schizophrenics are below the cutting score of 4.130 cycles per second, the larger part of the paranoid *S*s tested fall into the higher quartile ranges. If we wager that all scores below the median belong to schizophrenics and all high scores to paranoids, then we shall be correct in 33 cases out of 42. If we make the same wager for the normal-schizophrenic grouping, we shall be correct in 25 cases out of 31. The associated percentages of correct prediction are 78.5% for the schizophrenic-paranoid group and 80% for the schizophrenic-normal group where 50% correct predictions can be made by change alone.

Table 2

Significance of Mean and Rank Differences

Groups compared	*t*	Significance	x/sigma	Significance
Schizophrenic-Control	3.67	.001	3.80	.00009
Paranoid-Control	NS		NS	
Schizophrenic-Paranoid	4.35	.001	3.91	.00005

Discussion

The primary hypothesis of a difference in apparent motion perception between normal and undifferentiated schizophrenics seems to be tenable. The fact that these differences are more marked in this experiment than in the earlier experiment (5) may be due to the increase in sample size. The combined findings from the two experiments suggest that there is a noticeable deficit in the perception of optimal movement by the schizophrenic.

Since the results of the experimental procedure do agree reasonably well with the presence or absence of schizophrenia as established by other diagnostic criteria, the use of apparent motion perception as an added diagnostic procedure seems warranted. At the same time, due to the use of relatively complex electronic test equipment, the method seems more useful in connection with research procedures than as a routine diagnostic test.

The evidence concerning the "normal" perception of apparent movement in the paranoid schizophrenic is of considerable interest and

Table 3

Distribution by Cutting Scores

Cutting score	Schizophrenics	Paranoids	Controls	*N*
2.50–3.375	10	4	0	14
3.38–4.125	8	2	3	13
4.130–4.875	3	5	2	10
4.88–12.00	0	10	5	15
Total	21	21	10	52

merits separate discussion. The differences between schizophrenics and paranoid schizophrenics are so great as to suggest the presence of an entirely different pathology. It has been argued earlier (5) that the ability to perceive apparent movement is an index of cortical efficiency. If this is so, then the paranoid schizophrenic shows little evidence of lowered cortical efficiency in spite of the fact that as a group the paranoid schizophrenics tested showed considerable behavioral pathology.

In a first approximation to a theory of paranoid schizophrenia, it may be assumed that there are certain individuals who are genetically equipped with nervous systems which are highly efficient in the organization of perceptual processes. This assumption is not entirely groundless, since it is also implied in a genetic theory of intelligence.

Such an individual might be expected to go beyond the norm in organizing his perceptions into a coherent system. If there are severe personality problems present, this structure may become contaminated by personal reference or bias. Since the bias introduces some distortion of reality, it may be expected to change the individual's behavior towards the external world. This in turn begins to produce reactions on the part of society which must be incorporated into the meaningful if paranoid structure of perceptual organizations. The entire process may become circular in nature with a high degree of positive feedback. The accentuated emotional loading on the cortical processes will in time produce a schizoid behavior pattern.

Thus, the first stages of the pathology would be characterized by hyper-alert attitudes and complex perceptual organizations which would later be replaced by increasing loss of contact with reality and finally by random pathological behavior of a generalized hebephrenic nature. The presence of schizophrenic behavioral components in the paranoid individual may be viewed as a consequence of genetic accident plus the personal and private emotional problems acting together in a reactive rather than in a process fashion to produce such behavior.

It is likely that a long range study of paranoid schizophrenics would show that in the early stages of development they would show hypernormal ability to organize perceptual processes. In spite of behavioral pathology, they would not show "schizophrenic" deficit in this area. During the later course of the pathology, a decline in organizational ability would be accompanied by an increase in schizoid symptoms until finally the two pathologies could not be separated on a perceptual basis alone.

Summary

Earlier evidence has indicated that schizophrenics may show impaired ability to perceive visual apparent movement. This paper is a report of an experiment designed to explore the degree of impairment in a large sample of schizophrenics.

There are statistically significant differences between the undifferentiated schizophrenic group and the control group but not between the control group and the paranoid schizophrenic group. These findings support the theory of schizophrenic deficit outlined earlier by Saucer and Deabler (5) but require a new interpretation of the processes of paranoid schizophrenia. A tentative hypothesis concerning the nature of paranoid processes is discussed.

It is concluded that apparent visual movement thresholds may be a useful diagnostic tool but may be better suited to research than to routine diagnostic testing.

Received August 9, 1957.

References

1. Bender, M. B., & Teuber, H. L. Disturbances in visual perception following brain lesions. *J. Psychol.*, 1949, **28**, 223–233.
2. Gallese, A. J. Spiral afterimage as a test of organic brain damage. *J. clin. Psychol.*, 1956, **12**, 254.
3. Price, A. C., & Deabler, H. L. Diagnosis of organicity by means of the spiral afterimage. *J. consult. Psychol.*, 1956, **19**, 299.
4. Saucer, R. T. Processes of motion perception. *Science*, 1954, **20**, 806–807.
5. Saucer, R. T., & Deabler, H. L. The perception of apparent motion by organics and schizophrenics. *J. consult. Psychol.*, 1956, **20**, 385–389.
6. Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.
7. Werner, H., & Thuma, B. D. A deficiency in the perception of apparent motion in children with brain damage. *Amer. J. Psychol.*, 1942, **55**, 58–67.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

Success and Failure in Client-Centered Therapy As a Function of Client Personality Variables[1]

William L. Kirtner and Desmond S. Cartwright
University of Chicago

Cartwright (5) has presented evidence for the existence of a "failure-zone," ranging between 13 and 21 interviews, in client-centered therapy. In a subsequent paper, Taylor (25) presented closely comparable evidence with respect to psychoanalytically oriented psychotherapy.

Cartwright demonstrated that therapeutic outcomes were divided by the failure-zone and suggested that there may be two forms of therapeutic process, identifying these processes as "short" (1–12 interviews) and "long" (13–77 or more interviews). He further hypothesized (5, p. 362) "that certain individual differences between clients give rise to different kinds of therapeutic process."

On the basis of these results and hypotheses, five client groups were distinguished: short success, 1–12 interviews; short failure, 1–12 interviews; failure-zone, 13–21 interviews; long success, more than 21 interviews; and long failure, more than 21 interviews. Utilizing this schema, Kirtner (12) hypothesized that therapy length-by-outcome is related to the personality structures of clients at the beginning of therapy. The present paper reports an investigation of this hypothesis.

Method and Procedure

Subjects

From the research files of the Counseling Center, University of Chicago, all clients were selected for whom the following four conditions held: (*a*) the therapists' rating of outcome on a 9-point success rating scale (22) fell either between 7 and 9 inclusive ("success") or between 1 and 4 inclusive ("failure"); (*b*) the exact number of interviews was recorded; (*c*) a pretherapy Thematic Apperception Test (TAT) protocol was available; and (*d*) a recorded or transcribed first therapy interview was available. Twenty-four clients fulfilled all conditions. Two clients who fulfilled conditions (*a*), (*b*), and (*c*), but not (*d*), were also included, yielding $n = 26$ for this study. All clients were seen by experienced therapists (25 male and 1 female) between 1949 and 1954. The sample included 12 males, 14 females; 10 students, 16 nonstudents; the mean age was 27.5 years and the range 19 to 41 years.

Instruments

Rating scales were developed to measure twelve personality variables. These variables were selected to comprehend: (*a*) some of those variables widely referred to as designating level or pattern of personality integration, e.g., sex role and identification, motivation, impulse life; (*b*) some of those variables particularly relevant to the phenomenological basis of client-centered therapy, e.g., decision-making, rule usage, sense of self as a causative agent (16; 17, pp. 4–5); and (*c*) some of those variables especially associated with previous research in client-centered therapy, e.g., sensitivity to others (23, 24, 26), intrapunitive feelings (9). Upon analysis of the results derived from the application of the 12 scales, 6 scales were eliminated due to functional overlap with the remainder. Numerical ratings made by an independent judge on the 12 scales were intercorrelated. The resulting matrix was inspected for clusters from each of which one scale was selected, yielding four relatively independent scales, numbered I, IV, X, and XI. Scale VIII, while substantially correlated with Scale I, was uncorrelated with the other three selected scales and is therefore reported upon. Scale XII, highly correlated with Scale XI on numerical ratings, is also reported upon because of its distinctive qualitative differentiation between clients. Over the six scales, the range of intercorrelations was − .24 to .77. The median correlation was .29.

---

[1] This investigation is based upon a dissertation submitted by the senior author to the University of Chicago in partial fulfillment of the requirements for the degree of master of arts. It was supported in part by research grants from the National Institute of Mental Health, National Institutes of Health, U. S. Public Health Service, and from the Ford Foundation.

Following are brief descriptions of the six scales reported in this study. In complete form, all are 5-point, fully anchored scales.[2] A rating of 1 indicates minimal disturbance (optimal adjustment), and a rating of 5 indicates maximal disturbance.

Scale I—Impulse Life

From free flow and expression of generative impulses and ideas, to extreme:

- *a.* fright about impulses but with expression;
- *b.* control or repression;
- *c.* distortion of impulses.

Scale IV—Sense of Capability

From a sense of felt capacity to cope adequately with situations, to a sense of helplessness and lack of internal resources to cope with situations.

Scale VIII—Sex-role and Identification

From direct and open expression of sexuality with good control of feelings, to extreme conflict, uncertainty, and confusion about sex.

Scale X—People—Object Orientation

From the ability to relate deeply both to people and in activities, to:

- *a.* a desperate need for people;
- *b.* avoidance of relating to people with concomitant desperate need for activities;
- *c.* immobilization.

Scale XI—Sense of Comfort and Satisfaction

From a sense of self-comfort and satisfaction with the ability to locate problem sources both internally and externally in good balance, to:

- *a.* intensely driven to search himself for causes of discomfort and dissatisfaction;
- *b.* intensely driven to search outward for causes of discomfort and dissatisfaction;
- *c.* intense hopelessness to locate causes.

Scale XII—Punitive Feelings

From a self-appraising person who is not self-reproachful, to:

- *a.* extremely self-punishing person;
- *b.* extreme anticipation of punishment from external source;
- *c.* vacillation between "*a*" and "*b*."

It will be noted that some of the scales include alternate descriptive anchors, indicated by letters *a*, *b*, *c*. Such letter ratings occur only in connection with numerical ratings of 3 or more. As the level of disturbance increases in a given area (as the numerical rating is higher on these scales), the behavior displayed may take on one of several modal characteristics. Salient nonreactive, explosive, or other features of behavior become usual and representatively distinguishing attributes of the person rated. Such salient characteristics or modal patterns of behavior must be discretely described in order to obtain a fully anchored scale of levels of disturbance. As the level of disturbance decreases (as the numerical rating is lower on these scales), the behaviors displayed do not require multiple discrete descriptions; salient nonreactive, explosive, or other features of behavior are neither usual nor representatively distinguishing attributes of the person rated.

While the foregoing paragraph presents the descriptive (i.e., scale-making) problem, the existence of the problem invites explanation. The authors hold the working hypothesis that the individual adopts certain specifiable modes or techniques of behavior in order to control his internal disturbance or anxiety with respect to a given segment of his experience. As anxiety or disturbance intensifies, the techniques required for control of the disturbance, or for maintenance of personality organization, become more drastic and more representatively typical of the individual.

Ratings

Ratings of each client, on each of the six scales, were made on the basis of the pretherapy TAT protocol in conjunction with the recorded or transcribed first therapy interview. In two cases, the recorded first interview was not available, and the TAT protocol alone was used.

Ratings were made independently by the

---

[2] Copies of the complete scales and of the raw data from both judges may be obtained without charge from William L. Kirtner, Counseling Center, University of Chicago, 5737 Drexel Avenue, Chicago 37, Illinois, or for a fee from the American Documentation Institute. Order Document No. 5585, remitting $1.75 for microfilm or $2.50 for photocopies.

senior author and by another judge. The latter ("Judge II") knew neither the length nor the success rating of any case.

Reliability

Reliability of the therapist rating of outcome on the 9-point scale has been previously reported (5, 22) as slightly better than $r = .80$. To determine reliability for the numerical ratings on the six scales, all ratings by the two independent judges were correlated using Kendall's tau for ties in both rankings (10). The computational procedure employed was that developed by Cartwright (3). Corrections for continuity were made using Schaeffer and Levitt's (21) modification of Whitfield's correction (27). The routine for computing corrections for continuity and p values was that described by Cartwright (4). Table 1 presents these reliability data, which were considered adequate throughout.

Letter ratings occurred on four out of the six scales. Since such ratings are associated only with numerical ratings of 3 or more on these scales, a long and tedious problem in combinatorics would have to be solved to obtain an exact estimate of reliability, which did not seem justifiable. A fortiori, it did not seem justifiable to work on a solution for the reliability of joint numerical-plus-letter ratings. Instead, simple percentages of agreement were computed for all cases where both judges employed a letter rating, and regardless of numerical rating. Table 2 presents these data and includes, for the reader's information, the number of cases where neither judge used a letter rating.

No statistical analysis of results from letter ratings was made. However, the authors judge the data of Table 2 to warrant descriptive characterization of modal behaviors displayed by subjects rated on the four scales referred to.

Table 1

Reliability Coefficients (Tau) for Numerical Ratings on Six Scales

Scale	Tau	p
I. Impulse life	.65	.01
IV. Sense of capability	.62	.01
VIII. Sex-role and identification	.53	.01
X. People-object orientation	.73	.01
XI. Sense of comfort and satisfaction	.85	.01
XII. Punitive feelings	.83	.01

Table 2

Percentages of Agreement for Letter Ratings on Four Scales

	Number of cases for which letter ratings were used by		Agreements	
Scale	Neither judge	Both judges	Number	%
I.	3	22	20	91
X.	0	26	23	88
XI.	0	26	25	96
XII.	1	25	22	88

Analysis

The sample of 26 cases was divided into length-by-outcome groups according to the schema given. Table 3 summarizes descriptions of the resulting groups.

Regrettably for the research, only one long-failure case was among those available. This group was therefore eliminated from subsequent analysis. The remaining groups were examined in all possible pairs on each of the six scales as numerically rated by Judge II. All comparisons were made using tau for a dichotomy in one ranking as described by Kendall (10, pp. 32–34). The significance of differences between groups was computed in the manner previously indicated for the reliability analysis. In order to show the direction of differences between pairs of groups, mean values for each group on each scale were also computed.

Results

Numerical Ratings

Table 4 presents means for groups on each of the six scales and the p values of the differences between groups.[3]

Rejecting the null hypothesis if $p < .05$, each success group shows a lesser degree of

---

[3] See Footnote 2.

Table 3

Summary Description of Groups

Group	Number of interviews	Range of outcome ratings	N	Male	Female	Student	Non-student
Short failures	<13	1–4	10	4	6	4	6
Short success	<13	7–9	4	0	4	3	1
Failure-zone	13–21	1–4	7	5	2	1	6
Long failure	>21	1–4	1	1	0	1	0
Long success	>21	7–9	4	2	2	2	2

disturbance than either failure group on Scale I. Also, the short-success group shows a lesser degree of disturbance than either failure group on Scale VIII.

Of special interest are results setting a given group apart from the other three groups. We reject sets of three null hypotheses if in each instance $p < .10$. On this basis, Table 4 contains the following results: the short-success group shows less disturbance than any other group on Scales I and VIII; the short-failure group shows more disturbance than any other group on Scale IV; and the failure-zone group shows more disturbance than any other group on Scale VIII.

Letter Ratings

Table 5 presents the number of letter ratings agreed upon by both judges for four scales with respect to the four groups.

The most interesting feature of Table 5 is that success groups appear to differ from failure groups on Scales X, XI, and XII. While there were no differences on these scales as numerically rated for degree of disturbance, differences in typical mode of controlling inner disturbance were observed by both judges.

Distinguishing Descriptions

The results and observations mentioned in connection with Tables 4 and 5 are given embodiment in descriptions distinguishing the groups. Descriptions employ the language of those scales, and only those scales, for which either statistically significant differences or clear qualitative observations have been made.

The short-success group shows a generally higher level of personality integration than that descriptive of other groups. In particular, they are rather open to their impulse life (I), and, while showing some distortion and uncertainty, they are a good deal less confused about their sex role than other groups (VIII).

Though similar to the short-success group in many respects, the long-success group shows a somewhat lower level of personality integration in general. In particular, they are rather anxious about their impulse life (I)

Table 4

Significance Levels of Differences Between Groups on Six Scales, and Means Indicating Direction of Differences

Scale	Means of Groups[b]				p of Differences[a]					
	SS	LS	SF	FZ	SS vs. LS	SS vs. SF	SS vs. FZ	LS vs. SF	LS vs. FZ	SF vs. FZ
I.	2.00	3.00	4.10	4.43	<.10	<.01	<.01	<.01	<.01	>.20
IV.	3.00	3.25	4.30	3.71	>.20	<.02	>.20	<.03	>.20	<.07
VIII.	2.50	3.75	3.80	4.57	<.07	<.02	<.02	>.20	<.10	<.02
X.	3.50	3.75	4.10	4.00	>.20	>.20	>.20	>.20	>.20	>.20
XI.	3.75	4.50	3.90	3.57	<.15	>.20	>.20	>.20	<.08	>.20
XII.	3.25	4.25	3.80	3.85	<.19	>.20	>.20	>.20	>.20	>.20

[a] Two-tailed tests throughout.

[b] SS—short success; LS—long success; SF—short failure; FZ—failure-zone.

Table 5

Letter Ratings Agreed on by Both Judges

	Short success			Long success			Short failure			Failure zone		
Scale	*a*	*b*	*c*	*a*	*b*	*c*	*a*	*b*	*c*	*a*	*b*	*c*
I.				4			2	3	4		5	1
X.	4			4			1	4	4	2	1	2
XI.	3			4				8	2		6	1
XII.	3			4				8	1		4	1

and rather confused about the meaning and place of sex in their identity (VIII).

Both success groups differ from each failure group in showing less disturbance in their impulse life (I). Letter ratings show the success groups typically displaying a strong need for people to relate to in order to feel worthwhile (X), seeking internally for the cause and resolution of felt discomfort (XI), and rather strongly condemning, berating, and disvaluing themselves (XII).

The short-failure group shows a generally low level of personality integration. Especially distinguishing is their extreme underlying sense of incapacity to deal with their life situations (IV).

The failure-zone group also shows a generally low level of personality integration. Especially distinguishing is their extreme conflictedness about and rejection of their sex role (VIII).

Both failure groups show greater disturbance than either success group on several scales. They are extremely disturbed in their impulse life (I), with expressions frequently occurring in over-controlled or distorted modalities, as indicated by letter ratings. Overcontrol seems somewhat more typical for the failure-zone group. Letter ratings show the failure groups typically looking outward for the cause and resolution of their discomfort (XI), and typically anticipating punishment from external sources (XII).

Discussion

The foregoing results seem unambiguously consistent with the general hypothesis motivating the study: that therapy length-by-outcome is related to the personality structure of clients at the beginning of therapy. While caution must be invoked due to the small sizes of the groups studied, the results add to the growing body of knowledge about factors which favor or limit therapeutic change in client-centered therapy. Earlier work in this area has been summarized by Rogers (17, pp. 423–424, 427–428).

It may be noted that the present approach constitutes a synthesis of two trends in previous work on the general problem. One trend has dealt with the relations between pretherapy characteristics of clients and eventual ratings of improvement or outcome (e.g., 1, 2, 8, 11, 15, 19, 26). The other trend has dealt with the relations between pretherapy characteristics of clients and their length of stay in treatment (e.g., 7, 13, 14, 18, 20). To our knowledge, only one previous investigator, Dana (6), has attempted to predict improvement and length of treatment jointly. Due to the findings of Cartwright (5) and Taylor (25), a fusion of the two main trends into systematic emphasis upon length-by-outcome groupings now seems necessary. More generally, it is necessary to predict both amount and rate of change constituting improvement. While something of a beginning has been made in this direction by Dana's results and by those of the present study, much more needs to be done.

Summary

Twenty-six clients, who had completed a course of client-centered therapy at the Counseling Center, University of Chicago, were divided into five groups on the basis of length-by-outcome:

Short success (1–12 interviews)—*N* equals 4
Short failure (1–12 interviews)—*N* equals 10
Failure-zone (13–21 interviews)—*N* equals 7
Long success (more than 21 interviews)—*N* equals 4
Long failure (more than 21 interviews)—*N* equals 1

Twelve personality variables were selected, and 5-point rating scales were constructed to describe both intensity and mode of behavior with respect to each of the 12 variables. These scales were applied independently by two judges to the pretherapy Thematic Apperception Test protocol in conjunction with

the recorded or transcribed first therapy interview in each case. Interjudge agreement was deemed adequate throughout.

The ratings made by an independent judge were intercorrelated, and the resulting matrix was inspected for clusters from each of which one scale was selected, yielding four relatively independent scales. Two further scales were reported upon because, though substantially correlated with one of the four selected scales, they are uncorrelated with the others and point up interesting and distinctive differences between client personality structure.

It was thus found that therapy length-by-outcome is related to the personality structure of clients at the beginning of therapy. The most marked differences found on these scales were those between success groups and the failure groups.

Received July 31, 1957.

References

1. Barron, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, **17**, 327–333.
2. Barron, F. Some test correlates of response to psychotherapy. *J. consult. Psychol.*, 1953, **17**, 235–241.
3. Cartwright, D. S. A computational procedure for tau correlation. *Psychometrika*, 1957, **22**, 97–104.
4. Cartwright, D. S. A note concerning Kendall's tau. *Psychol. Bull.*, 1957, **54**, 423–425.
5. Cartwright, D. S. Success in psychotherapy as a function of certain actuarial variables. *J. consult. Psychol.*, 1955, **19**, 357–363.
6. Dana, R. H. The effects of attitudes towards authority on psychotherapy. *J. clin. Psychol.*, 1954, **10**, 350–353.
7. Gallagher, J. J. A comparison of clients who continue with clients who discontinue client-centered therapy. In Psychotherapy Research Group (Ed.), *Group report of a program of research in psychotherapy*. State College, Pa.: Pennsylvania State College, 1953. Pp. 21–38.
8. Gallagher, J. J. Test indicators for therapy prognosis. *J. consult. Psychol.*, 1954, **18**, 409–413.
9. Haimowitz, Natalie R., & Haimowitz, M. L. Personality changes in client-centered therapy. In Werner Wolff (Ed.), *Success in psychotherapy*. New York: Grune & Stratton, 1952. Pp. 63–94.
10. Kendall, M. G. *Rank correlation methods*. London: Griffin, 1948.
11. Kirkner, F. J., Wisham, W. W., & Giedt, F. H. A report on the validity of the Rorschach Prognostic Rating Scale. *J. proj. Tech.*, 1953, **17**, 465–470.
12. Kirtner, W. L. Success and failure in client-centered therapy as a function of personality variables. Unpublished master's thesis, Univer. of Chicago, 1955.
13. Kotkov, B., & Meadow, A. Rorschach criteria for predicting continuation in individual psychotherapy. *J. consult. Psychol.*, 1953, **17**, 16–20.
14. Lutes, B. C. A factorial study of an attrition group in client-centered therapy. Unpublished master's thesis, Univer. of Chicago, 1952.
15. Mindess, H. Predicting patients' responses to psychotherapy: A preliminary study designed to investigate the validity of the Rorschach Prognostic Rating Scale. *J. proj. Tech.*, 1953, **17**, 327–334.
16. Rogers, C. R. *Client-centered therapy*. Boston: Houghton Mifflin, 1951.
17. Rogers, C. R., & Dymond, Rosalind F. (Eds.), *Psychotherapy and personality change*. Chicago: Univer. of Chicago Press, 1954.
18. Rogers, L. S., Krauss, Joanne, & Hammond, K. R. Predicting continuation in therapy by means of the Rorschach test. *J. consult. Psychol.*, 1951, **15**, 368–371.
19. Rosenberg, S. The relationship of certain personality factors to prognosis in psychotherapy. *J. clin. Psychol.*, 1954, **10**, 341–345.
20. Rubinstein, E. A., & Lorr, M. A comparison of terminators and remainers in outpatient psychotherapy. *J. clin. Psychol.*, 1956, **12**, 345–349.
21. Schaeffer, M. S., & Levitt, E. E. Concerning Kendall's tau, a nonparametric correlation coefficient. *Psychol. Bull.*, 1956, **53**, 338–346.
22. Seeman, J. Counselor judgments of therapeutic process and outcome. In C. R. Rogers & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. of Chicago Press, 1954. Pp. 99–108.
23. Sheerer, E. T. An analysis of the relationship between acceptance of and respect for self and acceptance of and respect for others in ten counseling cases. *J. consult. Psychol.*, 1949, **13**, 169–175.
24. Stock, D. An investigation into the interrelations between the self-concept and feelings directed toward other persons and groups. *J. consult. Psychol.*, 1949, **13**, 176–180.
25. Taylor, J. W. Relationship of success and length in psychotherapy. *J consult. Psychol.*, 1956, **20**, 332.
26. Tougas, R. R. Ethnocentrism as a limiting factor in verbal therapy. In C. R. Rogers & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. of Chicago Press, 1954. Pp. 196–214.
27. Whitfield, J. W. Rank correlation between two variables, one of which is ranked, the other dichotomous. *Biometrika*, 1947, **34**, 292–296.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

The Performance of Psychoanalytic and Client-Centered Therapists in an Initial Interview[1]

Hans H. Strupp[2]
School of Medicine, University of North Carolina

As a technique for producing changes in the personality structure of a patient, psychotherapy does not exist apart from the person of the therapist. While it is axiomatic that the therapeutic instrument comprises the therapist, his personality, theory, and techniques, little systematic attention has been paid by researchers to these variables.

In a sense, the therapist is a reader of imaginary dials who acts "therapeutically" upon the information indicated by these measuring instruments within himself. When supplied with certain information, say, the behavior of a patient, he will register this information in certain ways, and he will react to it in certain other ways. The therapist's theory, experience, and personality determine the kinds of dials he uses as well as their calibration. His therapeutic actions are their reflection.

In order to obtain empirical data on the calibration of the therapist as a clinical instrument, we can supply a certain identical input (patient behavior) and note (*a*) how this input is perceived and evaluated; (*b*) what deductions and inferences are drawn from it; and (*c*) how these processes are translated into therapeutic action. On the hypothesis that this testing procedure will disclose differences among therapists, one can investigate systematic influences which might produce the discrepancies. For example, therapists may use different sets of measuring instruments whose readings have no relation to each other. They may use similar instruments, but the calibration may differ. They may interpret differentially the meaning of the same pointer reading. Any one of these possibilities, plus many others, may result in divergent therapeutic communications.

This analogy is of course a gross oversimplification of the therapist's activity in the therapeutic situation, but it may help to make explicit the objectives of the larger investigation of which this paper forms a part. The purpose was to study the therapist's performance in a first interview and to ask: If a group of therapists observe the same patient, how do they perceive the clinical problem? What kinds of evaluations result from their perceptions? How are their perceptions and evaluations related to their communications? And, can differences in their performance be accounted for in terms of common underlying variables? Specifically, this paper is concerned with a comparison between psychotherapists (psychologists) following psychoanalytic principles and therapists following the client-centered theory of Rogers.

Method and Procedure

To study the performance of psychotherapists under comparable conditions, a sound film of a first interview was used as an experimental model. The film was produced by the Veterans Administration for training purposes and was entitled "A Clinical Picture of

---

[1] This paper was presented at the 1957 meeting of the Eastern Psychological Association.

This research, part of a larger project supported by a research grant (M-965) from the National Institute of Mental Health, Public Health Service, was conducted while the author was affiliated with The George Washington University, School of Medicine.

[2] The author is indebted to Rebecca E. Rieger for her assistance and valuable comments.

Claustrophobia."[3] It was described as an unrehearsed interview from which no significant portions had been deleted. The motion picture was adapted for experimental purposes by introducing interruptions at 28 preselected points identified by a uniform title "What would you do?" followed by a number from 1 to 28.

Synopsis. The patient, a middle-aged man, complains of intense anxiety and gastrointestinal symptoms which appear to be psychogenic. He relates an episode of a panic reaction on a crowded street car. Following the incident he found himself unable to go out of the house unless he was accompanied by his wife, an invalid of many years. As the interview progresses, he dwells on an intense conflict with his mother, a strong-willed woman, and speaks at length about his "bad luck" in all phases of his life. The father had died when the patient was young, leaving him with family responsibilities for which he was unprepared. He feels that "I had three strikes against me before I came to the plate." He speaks freely, and his anger and resentment against his mother, brother, and people in general build up. The patient appears hostile, demanding, and dependent; he verbalizes a feeling that the world owes him a living. While it is not clear that he had psychotherapy before, he makes mention of previous interviews. He speaks of having obtained some "insight," of which his rationalizations and projections give little evidence. The interview ends somewhat abruptly after he has made a passionate plea for help, couched in a demand for reassurance.

The interviewer was an inexperienced resident in psychiatry who essentially permitted the patient to tell his story. He frequently repeated a word or phrase of the patient's narrative but focused very little. At times, he asked informational questions which seemed to have the effect of diverting the patient from an emotionally charged topic. The interviewer's attitude may be described as benevolent but rather distant. He was passive, offered no reassurance and seemed ill at ease. No clues were given about the outcome of the interview.

The film was shown to groups of therapists who were asked to assume that they were interviewing the patient, and to record in writing their communications to the patient while the film sequence was interrupted. Each pause lasted 30 seconds, which provided most therapists ample but not excessive time to make their comments. Respondents were free to respond or to remain "silent." The experimental version of the film ran continuously and lasted approximately 50 minutes.

Following the film showing, therapists were requested to complete a comprehensive questionnaire on diagnostic impressions, treatment plans and goals, formulations of the patient's emotional dynamics, problems in treatment, estimates of the degree of anxiety, emotional maturity, social adjustment, prognosis, the respondent's attitude toward the patient, etc. In addition, there was a biographical information blank containing 24 questions about the therapist's training, theoretical orientation, and current therapeutic activities.

The experimental procedure usually required about two hours.

The Samples

Cooperation was obtained from psychiatric and psychological societies, hospitals, training centers, and therapeutic institutes. Data were collected in Washington, Baltimore, New York, Perry Point, and Chicago. The total sample comprised 235 psychotherapists, 134 of whom were members of the medical profession (analysts, psychiatrists, and residents in psychiatry); 17 were psychiatric social workers. This paper is concerned with the responses of 14 psychologists who described their theoretical orientation as unqualifiedly Rogerian (8 were affiliated with the Counseling Center of the University of Chicago). This group is compared to 64 psychologists who professed to follow psychoanalytic principles broadly defined. Actually, they selected a total of 93 "labels," which were distributed over 14 categories; 72% of the choices were accounted for by: psychoanalytic orientation (general or orthodox), Sullivan, and eclectic.[4]

Summary statistics on age, experience, and personal analysis are presented in Table 1. The following data shed some further light on the composition of the samples:

---

[3] The "Psychotherapeutic Interviewing Series" was directed by Jacob Finesinger and Florence Powdermaker. The film under discussion, Part V of the series, was filmed in 1952. I wish to express my appreciation to the Psychiatry and Neurology Service of the Veterans Administration for granting permission to use the film for experimental purposes.

[4] Grateful acknowledgment is made to all psychologists who gave generously of their time to make this study possible. Thanks are also due the following organizations: VA Benefits Office, Washington, D. C.; VA Hospital, Perry Point, Md.; Maryland Psychological Association; Postgraduate Center for Psychotherapy, New York, N. Y.; Counseling Center, University of Chicago.

Table 1

Summary Statistics of Age, Experience, Personal Analysis

	Sample A[a]			Sample R[b]		
Statistic	*N*[c]	Mean	*SD*	*N*[c]	Mean	*SD*
Age	65	35.6 yrs.	8.7	13	33.4 yrs.	6.4
Experience in therapy	62	3.3 yrs.	3.1	12	3.0 yrs.	2.4
Personal analysis	33	291 hrs.	198	9	142 hrs.	101

[a] Psychoanalytically oriented psychologists.
[b] Rogerian psychologists.
[c] Fluctuations reflect absence of complete information and/or not having had personal analysis.

Two-thirds of the psychoanalytically oriented psychologists (Sample A) held the Ph.D. degree, as compared with 36% of the psychologists who followed Rogers' theory (Sample R).

Sixty per cent of the members in both groups said they had a fair amount or a great deal of experience in therapy and considered themselves competent therapists. All members of Sample R said they practiced client-centered therapy exclusively. Sample A showed a wider distribution, but most replies fell within the broad category of psychoanalytically oriented psychotherapy. Most Rogerians indicated that they had received very little training in psychoanalytic principles and techniques; 69% of Sample A claimed "a fair amount" or "a great deal."

Sixty-four per cent of Sample R had received their personal analysis within the client-centered framework. The responses of members of Sample A who indicated that they either had completed or were in the process of undergoing personal analysis (more than half the group) suggested that their analyses were based upon psychoanalytic theory.

Approximately one-half of the members in both groups stated that they devoted more than 25% of their time to the practice of psychotherapy. Almost two-thirds of Sample R specified their therapy as being primarily with neurotic patients. Thirty-eight per cent of Sample A responded similarly, but another third said that their patients represented a wider range of personality disorders.

Sample A appeared to treat their "average patient" for longer periods than Sample R: 38% of the former group described the length of therapy as exceeding one year, as contrasted with 14% in the latter. On the other hand, Rogerians predominantly saw the "average patient" twice a week, whereas the modal frequency for Sample A was closer to one weekly session.

Results

Therapists' responses to the film, that is, communications to the patient, were quantified by a multidimensional system of analysis which has been described previously (11). Questionnaire responses to each item were coded and the frequencies tabulated. In view of the small size of Sample R, response categories within an item had to be combined. Typically, this resulted in 2 × 2 tables which were tested by means of chi square (corrected for continuity).[5]

Clinical Evaluations and Therapist Attitudes

Responses to the question of diagnosis revealed no significant differences. It seems noteworthy, however, that no member of Sample R chose such labels as psychopath, schizoid, paranoid, or character disorder, which were used by appreciable minorities of Sample A, as well as by psychiatrists whose response distributions have been presented elsewhere (12). Asked to state the patient's major defense mechanisms, Sample R named a relatively smaller number than Sample A (ratio of 1:3), but the distributions did not appear to differ significantly. Nor did the samples differ in terms of qualitative descriptions of the patient's emotional dynamics.

Evaluations of the degree of ego strength, anxiety, insight, social adjustment, and disturbance disclosed no statistically significant differences.

The members of Sample R judged the prognosis with therapy to be more favorable than the members of Sample A ($p = < .05$): no

---

[5] A table showing the response distributions of the two samples in percentages and giving probability values for $p = < .10$ has been deposited with the American Documentation Institute. Order Document No. 5587, remitting $1.25 for microfilm or $1.25 for photocopies.

Rogerian therapist considered it unfavorable. By contrast, an unfavorable prognosis was given by 40% of the therapists in Sample A, who were in closer agreement with the predictions made by psychiatrists.

A larger proportion of Rogerian therapists professed a positive attitude toward the patient than analytically oriented psychologists ($p = < .02$). Members of the University of Chicago Counseling Center unanimously described their attitude as positive. Those responding negatively often stated that their attitude was influenced by the patient's hostility, anger, and dependency; those responding positively stressed his personality assets or saw him as "a patient with a problem," a response often given with some empathy.

Plans, Goals, and Difficulties in Therapy

Sample R showed greater reluctance to set up therapeutic goals than Sample A ($p = < .001$), but the two samples did not differ significantly as to estimated length of treatment or time needed to accomplish symptom relief. It seems pertinent to note that psychoanalysts were less optimistic than either of the psychologist groups.

Concerning difficulties expected in treatment, Sample R therapists were significantly less likely to mention: the patient's parasitic attitudes, poor motivation for therapy, acting out, or the possibility of psychosis ($p = < .02$). Nor were countertransference or problems of technique mentioned with any degree of frequency by Sample R ($p = < .01$). One-half of Sample R anticipated countertransference problems with this patient, as compared with almost 80% in Sample A; twice as many Rogerians answered the question in the negative ($p = < .07$). No Rogerian mentioned areas in the patient's living which should be avoided, at least early in therapy, whereas numerous areas were mentioned by Sample A therapists ($p = < .001$). Similarly, Rogerians were less likely to name areas in the patient's living which should be singled out for focus in therapy ($p = < .001$).

Rogerians either declined to specify attitudes or behaviors which the therapist should encourage in therapy with his patient or stressed the expression of feelings. Sample A therapists were more likely to stress: a sense of responsibility, increased socialization, relating feelings and symptoms to interpersonal situations, etc. ($p = < .02$).

Conversely, Sample A therapists were more likely to discourage attitudes and behaviors, such as: intellectualization, obsessive ruminations, self-pity, self-depreciation, helplessness, refusal to accept responsibility, demanding attitudes, and acting out; Rogerians, by comparison, tended to say they would discourage nothing or leave it to the patient ($p = < .001$).

No member of the Rogerian group advocated strictness by the therapist, as contrasted with more than a third of the members of Sample A who considered strictness therapeutically desirable ($p = .02$). A question about therapist's activity or passivity did not differentiate the two groups, possibly because the terms of reference were not clearly defined. Free association was discouraged by an appreciable proportion of Sample A therapists, but not by Sample R ($p = < .06$).

A question was asked about clues in the patient's verbal or nonverbal behavior which might not be used immediately but kept in mind for possible later reference. All members of Sample A who answered the question (85%) mentioned one or more clues. Four Rogerians specifically stated "none" ($p = < .001$). Rogerians commented about the patient's feelings and attitudes but, in contrast to Sample A, paid less attention to such clues as: gestures, bodily movements, manner of speaking, the patient's past or present interpersonal relations with his mother, wife, brother, or father—with or without inference about their dynamic significance.

Descriptions of the course of the transference did not diverge significantly, but members of the two samples disagreed about approaches to the problem ($p = < .001$). Rogerians predominantly encouraged or fostered a "corrective emotional experience" by conveying respect or clarifying feelings. Members of Sample A, on the other hand, tended to advocate interpretation and other procedures, such as reassurance and firmness, in which the therapist uses his role to induce changes in attitudes and actions.

Responses to a question dealing with "acting out," an aspect of the transference problem, revealed a similar pattern: no statistically significant differences with respect to its occurrence, but divergence in terms of "handling." Here, too, Rogerians recommended understanding, clarification, and reflections; analytically oriented therapists preferred an interpretive approach ($p = < .001$).

Evaluation of and Attitude Toward the Film Therapist

A number of questions were included to study similarities or differences between the film therapist's and the respondent's approaches to interviewing. It was hoped that in this way it might be possible to obtain a clearer picture of the respondent's own procedure.

Sample R therapists were more definite in their assertion that they would have conducted the interview "in a very different manner" ($p = < .02$). They dissociated themselves from his approach and tended to evaluate his performance as "a very inadequate job"; the majority of therapists in Sample A considered his performance reasonably adequate ($p = < .01$).

Within each sample, there was notable disagreement as to whether the therapist had been too passive or given insufficient support, but the intersample comparisons were nonsignificant.

Rogerians and psychoanalytically oriented therapists disagreed very markedly about the amount of time they would have devoted to obtaining data on the patient's life history: 75% of Sample R stated that they would have spent somewhat less or considerably less time than the film therapist; almost one-half of Sample A said they would have devoted somewhat more or considerably more time ($p = < .001$). The transcript shows that about 10% of the therapist's interventions can be construed as case history questions in the usual sense. These questions were usually asked in the context of the patient's story, and attempted to clarify a time, place, or interpersonal event. On the other hand, the therapist frequently repeated a word or phrase in a questioning tone of voice, apparently to obtain further data on a feeling or attitude.

The two samples revealed a statistically highly significant difference in terms of their attitude toward the therapist. The majority of Rogerian therapists described their attitude as negative, whereas most analytically oriented therapists professed a positive or neutral attitude ($p = < .001$).

A supplemental question disclosed that the respondents' criticisms were directed at the film therapist's technique as well as at his person. Prominently mentioned by both samples were the following objections: Therapist was not in control of the situation, engaged in too much activity but not of the "right" kind or not at the "right" time, too much interference with patient's talk, too much "accenting" (repeating last word), wrong emphasis, no plan. Rogerians criticized him especially for failing to focus on the patient's feelings and attitudes, and for a lack of empathy.

Interviewing Techniques

Analyses of therapists' communications to the patient disclosed that Rogerians strongly favored the reflection-of-feeling technique, which comprised 67% of their responses. The comparable percentage for analytically oriented psychologists was 17. The technique preferred by Sample A consisted of exploratory questions (40%). The number of "silent" responses given by Sample R was significantly smaller than that of Sample A. Communications showing a high degree of inference or initiative were rare in Sample R; Sample A had 3 to 8 times as many responses in these ranges.

As might be expected, Sample R therapists adhered quite consistently to the reflection-of-feeling technique, even when the patient pleaded for a direct answer from the therapist. Many analytically oriented therapists responded to this plea with a direct answer.

Discussion

In interpreting these findings it must be kept in mind that the sample of Rogerian therapists was small. This deficiency was partly mitigated, however, by the fact that this group was compared with a considerably larger sample of analytically oriented therapists to whom they were roughly comparable in terms of training and level of experience. Furthermore, on a number of questionnaire items, Sample R revealed almost complete unanimity which undoubtedly reflected homogeneous training in a single setting (The University of Chicago Counseling Center); Sample A, on the other hand, was made up of therapists from a variety of backgrounds. It may be noted that psychoanalysts (members of the American Psychoanalytic Association), whose training might be considered reasonably homogeneous, varied more widely in their responses.

The results showed a number of pronounced differences in approach which highlighted discrepancies between therapists subscribing to psychoanalytic principles and the client-centered framework. Both positions are too well known to require exposition in this context. It seems appropriate, however, to quote what Rogers considers to be the distinctive emphases of client-centered therapy:

> A second distinctive aspect [the first is the emphasis on research] is the deep confidence in the capacity of the individual, which client-centered therapists have found to be justified by their experience. Rather than seeing man as essentially a destructive animal [the reference is evidently to the death instinct postulated by Freud late in his career, one of the most disputed and least accepted tenets in contemporary analytic thinking], with the primary problem of the control of his impulses, we have found man to be essentially constructive, if he can be released from his defensiveness. Our experience is that, to the extent that he can fully accept his experience and himself, he becomes satisfactorily self-controlled. Our view of the basic nature of man differs sharply from those of most therapeutic orientations, which believe that guidance and control must be arranged by the expert.
>
> A third point of difference relates to a difference in theory. The theory of therapy which is developing in client-centered therapy stresses the importance of the immediate moment of experience in the relationship, not the genetic causation of behavior in the client, nor the formulation of some theoretic picture in the mind of the therapist (9, p. 207).[6]

Rogers' conception of the "ideal" therapist is that of a participant in an interpersonal situation whose primary function is to empathize with and understand the patient within his (the patient's) frame of reference. From this vantage point, the therapist is imposing an external frame of reference if he attempts to diagnose, assess, or prognosticate. The same dictum would apply to hypotheses or hunches the therapist may have about the etiology of the patient's conflict or how best to help him through psychotherapy. By this definition, the therapist who explicitly or implicitly goes beyond the patient's phenomenological self-descriptions is not client-centered.

---

[6] It is difficult to see how the terms "guidance and control," in the sense that Rogers seems to use them, do in fairness characterize the mainstream of contemporary psychoanalytic thinking. A semantic confusion seems to arise from the conception that the psychotherapist who functions in the role of a professional expert *ipso facto* manipulates the patient or controls him, and that guidance in the educational sense is synonymous with authoritarianism. Compare, in this connection, the formulations of a prominent psychoanalyst regarding therapeutic goals and the part played by the therapist:

> Treatment, of course, is aimed at the solution of the patient's difficulties in living and the cure of his symptomatology. Ideally these therapeutic goals will be reached by the growth, maturation, and inner independence of the patient. . . . This goal will also be actualized by the development of his capacity for self-realization, his ability to form durable relationships of intimacy with others, and to give and accept mature love (3, p. 34).
>
> . . . security and inner independence of the authoritarian values attributed to the conventional requirements of our culture are indispensable for the therapist who wants to guide his patients successfully toward finding out about the degree of cultural adjustment which is adequate to their personal needs (3, p. 33).
>
> The psychiatrist's respect for his patients will also help him to safeguard against the previously mentioned mistake of assuming an attitude of personal "irrational authority" instead of listening and conducting therapy in the spirit of collaborative guidance. This irrational authoritarian behavior will be harmful not only because it interferes per se with the patient's tendency toward growth and maturation but also, and more important, because it constitutes a traumatic repetition of the authoritarian aspects of the cultural pattern of behavior in general and of the parental pattern in particular, to which most mental patients have been harmfully subjected in their past (3, p. 17).

In his third point of difference, Rogers asserts that the experiential element in the psychotherapeutic situation is more important than the formulation of some theoretic picture in the mind of the therapist. This statement carries with it the implication that the therapist cannot adequately empathize with the patient, thereby helping to actualize the immediate moment of experience, if he attempts to understand rationally what is going on. Rogers has elaborated on this alleged dilemma in another paper (8). I, for one, know of no evidence that the therapist's private thoughts about the nature of the patient's emotional problem, its severity, etiology, etc., in and of themselves preclude or interfere with empathy, nor that the exclusion of the therapist's rational processes assures it. I feel that any psychotherapist does more than empathize with his patient: whether he interprets a feeling or attitude in terms of current or past interpersonal relationships, or whether he selects a particular feeling or attitude for reflection, he seems to follow a design for action dictated by his theory.

Psychotherapy based on psychoanalytic principles does differ from other forms of psychotherapy in its emphasis on giving the patient a rational understanding of his interpersonal processes as they emerge in the therapeutic relationship. The emotional experience remains uppermost, but, in addition, the patient gains a conscious appreciation of his own contribution to his difficulties in interpersonal relations.

Certain objections voiced by Rogerian therapists regarding the experimental procedure are understandable from this point of view. While they cooperated willingly, they made it clear that ordinarily they do not work this way.

On this basis, too, the following results are consistent with the client-centered viewpoint: a reluctance to make a diagnosis beyond the general label "anxiety," to formulate therapeutic goals, to mention dangers that might be encountered in therapy or cautions that would have to be observed, to encourage or discourage specific attitudes or behaviors in the patient, to treat him with strictness, to discourage the free expression of feelings (free association), or generally to consider problems of technique or "handling." Rogerians were also disinclined to make inferences about the patient's behavior in the interview situation, the etiology of his disturbance, or the dynamic import of his behavior. Finally, they de-emphasized case history questions.

One of the sharpest discrepancies between client-centered and analytically oriented therapists related to the question of prognosis: *no* Rogerian gave an unfavorable estimate, whereas 40% of Sample A took a pessimistic view. How can this finding be explained?

First, it must be considered that our knowledge of neurotic disorders and psychotherapy is not sufficiently developed to permit specific predictions about the outcome of treatment even when a great deal is known about the patient. Such predictions undoubtedly become increasingly accurate as therapy progresses, and the therapist comes to learn more about the nature of the conflict, its ramifications, its history, and the patient's current level of functioning. Thus, a prognostic estimate based on a single interview may be little more than a guess, which may reveal more about the therapist than about the patient.

Second, the key issue in this comparison seems to be the high degree of uniformity among Rogerians which was not paralleled by the analytically oriented therapists whose responses appeared to be more "normally" distributed. If therapists were talking about a specific event, the predictions of some obviously must be right and the predictions of others must be wrong. Unquestionably, therapists were not talking about a specific event but about a therapeutic outcome concerning which a wide range of ideas and opinions were entertained.

In an attempt to shed some light on this result, I propose to examine the following alternatives: (*a*) Rogerians gave a more favorable prognosis because they perceived the patient as less disturbed; (*b*) Rogerians planned for more modest therapeutic goals and therefore gave a more optimistic prognostic estimate; (*c*) prognostic estimates may be relatively unrelated to the presenting problem or the therapist's past successes with similar patients, but reflect more directly the therapist's attitude, his theoretical orientation, or combinations of these.

(*a*) The available evidence indicates that therapists in both samples tended to regard the patient as seriously disturbed. They were also reasonably agreed that he suffered a great deal of anxiety, that he was emotionally immature, socially maladjusted, and that he had relatively little insight. There is no reason to believe, therefore, that Rogerians judged the patient's difficulties as less severe.

(*b*) It is difficult to judge similarities or differences in therapeutic goals because of the vagueness of the terms of reference. Rogerians tended to advocate greater self-acceptance; analytically oriented therapists tended to strive for some insight. However, the latter group generally did not recommend the more intensive forms of psychotherapy usually associated with psychoanalysis. On this basis, it seems fair to say that the Rogerians' more favorable prognostic estimates were not a function of more limited therapeutic goals.

(*c*) It has been reported previously (12) that there was a highly significant statistical relationship between psychiatrists' prognostic estimates and their self-ratings of their conscious attitude toward the patient ($r = .39$). For psychoanalytically oriented psychologists, the correlation was $r = .42$. Among Rogerians, the invariance between attitude and prognosis was almost complete. Eight therapists at the University of Chicago Counseling Center expressed a positive attitude and rated the prognosis favorable.

In the case of psychiatrists, it was found that both attitude and prognosis were significantly associated with other clinical evaluations and treatment plans. In addition, therapists' attitudes tended to be reflected in the emotional tone and the degree of empathy of their communications. These findings suggested the interpretation that clinical assessments tended to reflect, in part at least, the therapist's emotional reaction to the patent.[7] On the other hand, if the therapist had undergone personal analysis, his empathy tended to be greater irrespective of his conscious attitude. In other words, personal analysis seemed to introduce some distance between a negative reaction to the patient and the therapist's ability to empathize with him. It is possible, then, that the therapist's conscious attitude toward the patient is less important than what he does about it.[8]

To integrate a workable therapeutic relationship, the therapist, according to psychoanalytic theory, must convey an attitude of acceptance, patience, respect, etc. He must also be able to empathize (1, p. 580). This does not mean that he must have or should have a positive attitude toward *all* aspects of the patient's behavior. Rather, it is recognized that the therapist, like other people, may experience emotional reactions. However, he is enjoined to be vigilant to make sure that they do not interfere with the therapeutic process and if they do, to seek clarification through self-analysis or other means.

The emphasis on diagnosis and etiology, according to analytic theory, determines at least partially the therapist's approach to therapy. Reich says:

> It should be clear that one approaches an aggressive patient unlike a masochistic one, a hyperactive hysteric unlike a depressive one, that one changes one's attitude to one and the same patient according to the situation, that, in brief, one does not behave neurotically oneself, even though one may have to deal with some neurotic difficulties in oneself.
>
> One cannot give up one's own individuality, a fact which one will consider in the choice of patients. But one should be able to expect that this individuality is not a disturbing factor and that the training analysis should establish the necessary minimum in plasticity of character (6, p. 139).

From these considerations it would follow that, ideally, the prognosis for any one patient represents a realistic assessment of the severity of the disturbance reflecting clinical experience regarding the response of similar patients to psychotherapy.

While evidence has been presented (12) that a prognostic estimate *may* be based on the therapist's emotional reaction to the patient rather than on the clinical evidence, it does not follow that an unfavorable prognosis is *ipso facto* proof of a rejecting attitude on the therapist's part. It may well be a realistic assessment of the clinical picture.

It appears that the results obtained from client-centered therapists must be explained on different grounds. First, one of the most important conditions for constructive personality change postulated by Rogers (10) is the therapist's "unconditional positive regard" for the patient. This means a "prizing" of the person, the absence of *conditions* of acceptance and of a selective evaluating attitude. Unconditional positive regard appears closely related to empathy, another of Rogers' key postulates, which is defined as the therapist's "experiencing an accurate, empathic understanding of the client's awareness of his own experience" (10, p. 99).

---

[7] It seems reasonable to assume that clinical evaluations, treatment plans, and goals are a function of the diagnosis. Comparisons of therapists (based on the total sample) giving a "pure" diagnosis of Hysteria, Obsessive-compulsive, Paranoid, and Anxiety showed a number of statistically significant differences in terms of clinical evaluations, etc. (12). It was pointed out in the same connection that the choice of certain diagnostic labels (paranoid, psychopath) may signify a subtle value judgment which in turn may be determined by the therapist's attitude toward the patient. The data seemed to favor this conclusion.

[8] It should be recalled that the operational definition of therapist's conscious attitude toward the patient, as used in this investigation, was a self-rating on a five-point scale. Apart from considerations of reliability, this rating was undoubtedly influenced by therapists' willingness to express their genuine conscious attitude, unconscious distortions, etc. It is difficult to say to what extent such ratings represent countertransference reactions even if there were greater agreement about the term (5). It seems that Gitelson's (4) description of reactions to the patient as a whole, which he terms transferences of the analyst as distinguished from reactions to partial aspects of the patient (countertransference), fits the present ratings reasonably well:

> It is my impression that total reactions to a patient are *transferences* of the analyst to his patients and are revivals of ancient transference potentials. These may be manifested in the over-all attitude toward patients as a class or may exacerbate in the "whole response" to particular patients. These attitudes may be positively or negatively toned. They are likely to manifest themselves early in the contact with a patient and determine the tendency of the analyst towards the whole case (4, p. 4).

Second, in explicitly rejecting the usefulness of diagnosis for psychotherapy, Rogers denies that the therapist's approach or operations should be influenced or modified by the type of patient or disturbance (10, p. 101). This would also imply that prognostic estimates, to the extent that they attempt to take into account the severity or chronicity of the patient's disturbance, are pointless.

Thus, it appears logical that the client-centered therapist who fully accepts the theory must (*a*) describe his attitude toward the patient as positive and (*b*) give a favorable prognosis. The ratings under discussion seem to reflect very clearly the respondents' theoretical orientation and are perhaps best explained on this basis.

Whether a conscious attitude toward the patient is a more or less individual expression of the therapist's emotional reaction or whether the attitude is an intrinsic part of the therapist's theoretical framework, or both, the more important question relates to the manner in which this attitude is expressed in the therapeutic situation. Some evidence (12) has been adduced that there is in fact a carry-over. It remains to be demonstrated to what extent the therapist's initial attitude toward the patient influences the course and outcome of therapy, but there seems to be a very real possibility that the therapist making a negative evaluation and/or reacting negatively to the patient may unwittingly contribute to the realization of the events he diagnoses or predicts. Contrariwise, the therapist who brings a genuinely positive approach to bear upon the relationship may thereby promote a more favorable outcome.[9] Both psychoanalytic and client-centered theory emphasize that the person of the therapist, as revealed in this relationship, will in time serve as a new, more mature, and more desirable model of reality than past interpersonal relationships which have distorted the patient's perceptions of himself and others (13):

> What is curative in the [therapeutic] process is that in tending to reconstruct with the analyst that atmosphere which obtained in childhood, the patient actually achieves something new. He discovers that part of himself which had to be repressed at the time of the original experience. He can only do this in an interpersonal relationship with the analyst, which is suitable to such a rediscovery. . . . Thus, the transference phenomenon is used so that the patient will completely reexperience the original frames of reference, and himself within those frames, in a truly different relationship with the analyst, to the end that he can discover the invalidity of his conclusions about himself and others (7, p. 151).

Rogers puts it this way:

> In a minimal way, the client may perceive from the first the unconditional positive regard of the therapist for him. The perception is continually strengthened as he discovers that each facet of himself which is exposed—contradictions, weaknesses, strengths, abnormal feelings, tender feelings, vicious attitudes, antisocial behaviors, fears and despairs—are all met with equal positive regard, because each of these elements is a part of him, and he is prized unconditionally. The client then gradually takes in the experience of being loved—a love which is nonpossessive and nondemanding, warm but not oversolicitous. As he lives for a sequence of hours in this atmosphere, the experience has two primary results. It permits him to relax the tight defensive structure of his concept of himself and to admit into awareness and fully experience attitudes which previously he had found too threatening. . . . In the second place, the attitude of the therapist toward him is gradually internalized so that he can take the same attitude toward himself. He comes to prize himself, to feel that he *is* of value (9, p. 204).

Much remains to be learned about the therapist's personality and its contribution to the treatment process, but, as I have suggested elsewhere (11), this focus on the person of the therapist may be one of the most fruitful avenues for research in this area.

Summary

This investigation was designed to obtain empirical evidence on psychotherapists' performance in an initial interview for the purpose of studying the influence of relevant therapist variables.

Psychotherapists were asked to respond as vicarious interviewers to a sound film of an unrehearsed initial interview. Data were collected from 235 psychotherapists in several cities. The present paper, part of a larger project, was concerned with a comparison between 64 psychoanalytically oriented therapists and 14 therapists following Rogers' client-centered theory. All respondents were psychologists, and 8 Rogerians were staff members of the University of Chicago Counseling Center.

---

[9] If it turns out that the therapist's personality—his dedication, faith, belief—is indeed a primary force in modern psychotherapy, painstaking research may still give a clearer picture of the components. (See in this connection [2].)

Statistical comparisons of questionnaire responses to a series of items dealing with diagnosis, prognosis, treatment plans, and attitudes disclosed a number of significant differences between the two samples. Rogerians were generally disinclined to make plans for treatment or to set up therapeutic goals. Their prognostic estimates were more favorable, and they rated their attitude toward the patient as more positive than analytically oriented therapists.

In discussing the implications of these findings, attention has been drawn to the interrelationships between therapist's attitude toward the patient and clinical evaluations, including diagnosis and prognosis. In view of the self-realizing character of one's expectations, the therapist's attitude toward the patient may have important effects upon the course and outcome of psychotherapy.

This investigation has stressed the great need for research on the therapist's contribution to the treatment process.

Received August 9, 1957.

References

1. Fenichel, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.
2. Frank, J. D. Some effects of expectancy and influence in psychotherapy. In *Progress in psychotherapy.* Vol. 3. New York: Grune & Stratton, 1958. Pp. 27–43.
3. Fromm-Reichmann, Frieda. *Principles of intensive psychotherapy.* Chicago: Univer. Chicago Press, 1950.
4. Gitelson, M. The emotional position of the analyst in the psychoanalytic situation. *Int. J. Psychoanal.,* 1952, **33**, 1–10.
5. Orr, D. W. Transference and countertransference: A historical survey. *J. Amer. psychoanal. Ass.,* 1954, **2**, 621–670.
6. Reich, W. *Character-analysis.* (3rd ed.) New York: Orgone Inst. Press, 1949.
7. Rioch, Janet. The transference phenomenon in psychoanalytic therapy. *Psychiatry,* 1943, **6**, 147–156.
8. Rogers, C. R. Persons or science: A philosophical question. *Amer. Psychologist,* 1955, **10**, 267–278.
9. Rogers, C. R. Client-centered therapy: A current view. In Frieda Fromm-Reichmann & J. L. Moreno (Eds.), *Progress in psychotherapy 1956.* New York: Grune & Stratton, 1956. Pp. 199–209.
10. Rogers, C. R. The necessary and sufficient conditions of therapeutic personality change. *J. consult. Psychol.,* 1957, **21**, 95–103.
11. Strupp, H. H. A multidimensional system for analyzing psychotherapeutic techniques. *Psychiatry,* 1957, **20**, 293–306.
12. Strupp, H. H. The psychotherapist's contribution to the treatment process. *Behav. Sci.,* 1958, **3**, 34–67.
13. Strupp, H. H. Toward an analysis of the therapist's contribution to the treatment process. *Psychiatry,* in press.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

A Theoretical Framework for Nondirective Play Therapy: Concepts from Psychoanalysis and Learning Theory

Dell Lebo

Richmond Professional Institute

As Freud's systematic thinking progressed, aggression emerged from a position of relative obscurity in 1905 into one of decided prominence in 1930. At first, aggressive impulses were regarded as derivatives or components of sexuality (10). Later Freud contended that each instinct had its own power of becoming aggressive, deprivations being especially productive of aggressive responses (9).

Freud's formulation of the death instinct (6, 7) next led him to assert that a portion of that instinct, directed toward the outer world, was manifested as an aggressive instinct. However, it was not until 1930 (8) that a substantial body grew onto the theoretical framework of aggression. A desire for aggression was then said to be an independent, innate drive. This instinctual disposition varied in intensity from man to man. Libido and aggression were viewed as primary instincts entwined with strivings toward life and death (6).

The relationship between aggression and deprivation was experimentally examined by learning theorists and found to be relevant. Consequently, learning theory has given attention to aggression. At first it was assumed that frustration was a direct antecedent of aggression (2). As the investigations proceeded, it was realized that frustration could give rise to a variety of responses, one of which could be some form of aggression (17). Aggression was viewed as a secondary, rather than as a primary, drive because it did not seem to involve any biological changes in the organism. Instead, aggression was held to be a simple, learnable response, subject to modification. The selective agent in the learning of adult aggressive habits was reinforcement. The position of aggression in the hierarchy of responses to frustration depended upon whether past conditions were such as to reinforce aggression as a response to external cues. Despite a continued emphasis on the importance of learning in aggression, learning theorists no longer deny the role of innate factors (3).

There is an interesting parellel here. Two powerful psychological viewpoints seem to have gradually admitted aggression into more and more of their hypothetical realm. Such a similar change in two independent systems of thought may foreshadow the eventual importance of aggression in other psychotherapeutic theories. The theoretical development of aggression indicates it would be worthwhile to consider its possible importance to a psychotherapeutic theory that does not at present include it.

Aggression and Client-Centered Theory

One of the newest of psychological systems, the nondirective or client-centered approach, has not directly incorporated aggression. The fourth proposition of Roger's formulation of personality is concerned with "the tendency of the organism to behave defensively in the face of threat, to achieve the goal of self-maintenance even when the usual pathway to that goal is blocked" (22, p. 488). However, he included these tendencies, which might sound like aggression, with food assimilation and a general striving toward maturation, the tendency to which all organic and psychological needs might be ascribed by nondirectivists. If the importance achieved by aggression in the two concepts previously discussed is a proper forecast, then it appears likely that it is through this proposition that aggression will enter nondirective personality theory.

Age and Aggression

In both psychoanalytic and learning theory the appearance and expression of aggression are closely dependent upon both maturational and experiential factors which follow a fairly close chronological sequence. To psychoanalysts, genetic factors must be considered in the establishment of libidinal development and characteristic defenses. It would be clinically incorrect for them to assume exclusively that an aggressive child is reacting constantly to external frustration and to exclude the considerations of inner sources such as ego disturbances and libidinal factors.

In analytic work with children, aggressive qualities reach their peak with the anal stage and become especially evident also during the Oedipal period (5). When the child reaches the stage in which his ego or superego disapproves of aggression turned outward, aggressive impulses will be directed to substitute objects, or modified by restriction, sublimation, and internalization (12).

The learning theorists explain these changes on another basis. Both schools would probably agree that the child is better able to handle instinctual drives by the acquisition of language. Psychoanalysts, however, see language integration as enabling ego mastery to proceed, while the learning theorists see it as an example of further learning which enables earlier, more diffuse, responses to be eliminated and more selective responses, such as verbal aggression, to be utilized. This verbal aggression, however, is subject to modification by reinforcement. As the child grows older and reactions with adults become important to him, his pattern of aggression is modified. Progression from infantile diffusion to directed attacks, to verbal specialization, to a marked reduction seems to be the behavioral pattern of aggression expression in normal children.

As a child grows chronologically, certain changes occur which influence the pattern of aggression. Whether these changes are called a developing superego or a modification through learning experiences, the position of aggression in the hierarchy of available responses does seem to change in an observable manner.

Other important behavioral changes take place from one age level to another. Gesell and Ilg have indicated that as a child grows older the "interior patterns of his private psychology" are changing as well (11, p. 12). Such changes have been a concern of psychoanalytically oriented play therapists who advocate different treatment approaches to children of different ages (16).

Age in Nondirective Play Therapy

Just as the concept of aggression has not been admitted into client-centered philosophy, so has the concept of personality changes with advancing age been lacking from formulations of the nondirective play therapy process. Perhaps the conceptualization of nondirective play therapy (1) excluded consideration of age differences because of the emphasis on *process* in nondirective therapy in play therapy: "Experiences with play therapy . . . add to the conviction that therapy is very definitely a process, a process which follows certain fundamental trends" (21, p. 170).

Nondirective play therapy, based directly upon the techniques of nondirective counseling, emphasizes the forces within each individual to find mature behavior more satisfying than immature behavior. "This type of therapy starts where the individual is and bases the process on the present configuration . . ." (1, p. 15). Since the individual is accepted where he *is* in client-centered therapy and the process begun, the child is accepted similarly where he is chronologically and the process of nondirective play therapy begun.

There may be a recognition that personality changes from age to age in the expression of the one basic nondirective need, namely, the striving for maturity. However, such an implied recognition of age differences does not enter into the method of nondirective play therapy. The process is the same for all children admitted to the playroom regardless of age.

Studies of the Process of Nondirective Play Therapy

As has been shown, both age and aggression have been found to be factors of keen importance in other theories of psychotherapy, especially where children were concerned. Yet neither factor has been considered generally in studies of the process of nondirective play therapy. Such studies fall into two types, methodologically as well as chronologically distinguishable.

Studies of the earlier type involved examination of the process of nondirective play therapy as found in already available protocols. Later studies were objective investigations of the protocols of selected children.

Protocol-Centered Studies

Landisberg and Snyder (13) examined the protocols of four cases to determine what took place in nondirective play therapy. They found similarities in the process from case to case. Although the four children ranged in age from five to six years, the categories used to classify the verbal remarks had been developed for use with adults. Since the categories did not originate from children's verbalizations, their representativeness may have been forced and unrealistic. Furthermore, the narrow age range prevented inferences regarding the process of nondirective play therapy at different age levels.

The Procrustean forcing of children's statements into adult categories was obviated by Finke (4) who developed new categories based on an analysis of children's statements made in nondirective play therapy. She used protocols from six children ranging in age from five to eleven years. It was found that the children, though in therapy with different therapists, showed similar trends in their verbal expressiveness. Her study did not consider the possibility that just as the wide age discrepancies between adults and children might influence their verbalizations, so might children's categories vary significantly from one age level of maturity to another.

Seemingly the last study of this type was submitted for publication in 1952 (18). It was concerned with similarities in the pattern of emotional growth of normal and disturbed children. Extensive excerpts, arbitrarily selected and subjectively interpreted, revealed that "the emotional process and the changes in feeling tones evident as a child goes through play therapy . . . occur . . . not step by step, but in individually varying sequences" (18, pp. 97–98).

This study was nonobjective and resembled the suggestive investigations appearing in the persuasive or propagandistic period of research in nondirective play therapy (15). Nowhere was consideration given to age.

Child-Centered Studies

The first publication of the later type of investigation was apparently that of the present writer (14). Twenty children, aged 4, 6, 8, 10, and 12 years, equated for intelligence and social adjustment, were used as subjects. Maturation, as represented by chronological age, was found to account for different types of statements made in the nondirective play situation.

Similar studies by Moustakas (19, 20) have appeared since. One of them (19) studied the frequency and intensity of the expression of negative attitudes of nine well-adjusted and nine disturbed young children matched in various characteristics. At the end of four play therapy sessions, both groups expressed about the same types of negative attitudes.

Another example of subject-centered research concerned 10 four-year-old children with and without serious emotional problems (20). "Data obtained . . . showed more similar than divergent behavior in the two groups . . ." (20, p. 155). The most significant difference between the groups was that the maladjusted children were nonattentive. This difference was explained by the maladjusted children's noninteractive behavior, i.e., fantasy or activities that excluded the therapist. The degree of aggressiveness was not controlled. One wonders if the results would have been the same had aggression been considered.

The above mentioned studies represent all of the published work on the process of nondirective play therapy known to the present writer. While the studies are similar they are not directly comparable. In only four investigations does age seem to have been controlled. In three publications, age was regulated by restriction; only children within a narrow age range were studied: 5–6 years (13) and 4 years (19, 20) respectively. Hence, only one study clearly took age differences into account (14). This study alone found definite changes in the types of statements made by children in nondirective play therapy.

Only two of the studies clearly controlled intelligence (14, 19). Hence, differences in intellectual ability, long recognized as being of fundamental importance in speech habits (*vide* the items of the Stanford-Binet Intelligence Test), have not been considered in many studies of the verbalizations made in nondirective play therapy. This lack would seem to be of sufficient magnitude to vitiate investigations that have failed to control for it. None of the studies considered the possibility of differences arising in verbalizations as related to aggressiveness or lack of aggressiveness in children.

The empirical progress made in psychoanalysis and learning theory has been anticipated by the construction of adequate theory. Further developments in nondirective play therapy would seem to require suitable theory. Without such a theoretical framework, nondirective play therapy may be in danger of becoming immeshed in the "regularity" of its process, until, like introspectionism, it becomes process bound and unproductive. A good theory is said to be economical in permitting many observations to be subsumed under a single proposition. Theories concerning changes ascribable in aggression or age would seem to be both economical and meaningful.

Summary

Psychotherapeutic processes may vary considerably as a function of both the age of the subject and the dynamic processes characterizing his psychopathology. In childhood, the relationship between age and aggression is especially close. Nondirective play therapy does not formally consider the influence of aggression or age in its method. As a result, studies of the process of nondirective play therapy generally indicate that all children undergo the same process. It would seem, however, that if age and aggression were evaluated properly, the course of nondirective play therapy would vary with the age of the child and with the character of the psychodynamic processes, e.g., aggression, associated with or defining his problem. Insistence on the sameness of the process, and consequently of the method, may only result in the tautological death of productive work in nondirective play therapy.

These assertions were supported by inferences drawn from the history of Freudian and learning theory and by data from studies of the process of nondirective play therapy.

Received August 12, 1957.

References

1. Axline, Virginia M. *Play therapy.* Boston: Houghton Mifflin, 1947.
2. Dollard, J., Doob, L., Miller, N. E., Mowrer, O. H., & Sears, R. R. *Frustration and aggression.* New Haven: Yale Univer. Press, 1939.
3. Dollard, J., & Miller, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.
4. Finke, Helene. Changes in the expression of emotionalized attitudes in six cases of play therapy. Unpublished master's thesis, Univer. of Chicago, 1947.
5. Freud, Anna. Notes on aggression. In S. Lorand (Ed.), *The yearbook of psychoanalysis.* Vol. 6. New York: International Universities Press, 1951. Pp. 145–154.
6. Freud, S. *Beyond the pleasure principle.* London: Hogarth Press, 1922.
7. Freud, S. *The ego and the id.* London: Hogarth Press, 1927.
8. Freud, S. *Civilization and its discontents.* London: Hogarth Press, 1930.
9. Freud, S. *A general introduction to psychoanalysis.* New York: Liveright, 1935.
10. Freud, S. Three contributions to the theory of sex. In A. A. Brill (Ed.), *The basic writings of Sigmund Freud.* New York: Modern Library, 1938. Pp. 553–629.
11. Gesell, A., & Ilg, Frances L. *The child from five to ten.* New York: Harper, 1946.
12. Hartmann, H., Kris, E., & Lowenstein, R. M. Notes on the theory of aggression. In Phyllis Greenacre *et al.* (Eds.), *The psychoanalytic study of the child.* Vol. 3–4. New York: International Universities Press, 1949. Pp. 9–36.

13. Landisberg, Selma, & Snyder, W. U. Nondirective play therapy. *J. clin. Psychol.*, 1946, **2**, 203–213.
14. Lebo, D. The relationship of response categories in play therapy to chronological age. *J. Child Psychiat.*, 1952, **2**, 330–336.
15. Lebo, D. The present status of research on nondirective play therapy. *J. consult. Psychol.*, 1953, **17**, 177–183.
16. Lebo, D. Age and suitability for nondirective play therapy. *J. genet. Psychol.*, 1956, **89**, 231–238.
17. Miller, N. E., Sears, R. R., Mowrer, O. H., Doob, C. W., & Dollard, J. The frustration-aggression hypothesis. *Psychol. Rev.*, 1941, **48**, 337–342.
18. Moustakas, C. E. Emotional adjustment and the play therapy process. *J. genet. Psychol.*, 1955, **86**, 79–99.
19. Moustakas, C. E. The frequency and intensity of negative attitudes expressed in play therapy: A comparison of well-adjusted and disturbed young children. *J. genet. Psychol.*, 1955, **86**, 309–325.
20. Moustakas, C. E., & Schalock, H. D. An analysis of therapist-child interaction in play therapy. *Child Develpm.*, 1955, **26**, 143–157.
21. Rogers, C. R. *Counseling and psychotherapy*. Boston: Houghton Mifflin, 1942.
22. Rogers, C. R. *Client-centered therapy*. Boston: Houghton Mifflin, 1951.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

Personality Patterns of Patients Before and After Application for Psychotherapy[1]

Norman S. Greenfield
University of Wisconsin Medical School

A previous study (1) utilized the MMPIs of 132 male college students who later at some time in their four-year college course sought psychotherapy at a student health psychiatric clinic, with the aim of testing the hypothesis that promptness of referral is positively related to the degree of anxiety, the extent of internalization of problems, the amount of discomfort, and the severity of pathology. The results were surprising in that none of the variables proved to be significantly related to the number of months which intervened between the time of taking the test and appearance at the clinic. Thus, in the search for precise motivational variables influencing utilization of psychotherapeutic services, it was impossible to verify even such simple "truths" which have so long received consensual validation. It was concluded that for the population studied, at least, these measures were not indices of therapeutic readiness.

Having demonstrated an inability to predict which students would eventually become patients and which would not in a setting wherein psychotherapy was equally available to all students at no cost to the patient, the question was then raised as to what personality factors, or more precisely what changes in personality, influenced the decision to apply for treatment. Again, logic and clinical experience led us to assume that, other factors being equal, a person seeks psychotherapy at a time when he is feeling worse, or suffering more than when he did not seek psychotherapy.

The routine administration of the MMPI[2] to entering students made possible an examination of self-controlled profiles of 31 students who were given a second MMPI at the clinic. The *S*s were 20 males and 11 females whose mean age was 19 years. The mean number of months intervening between first and second test administrations was 10.97, with a range from 1 to 26 months.

Clinically, the presenting complaints of the patients as well as their psychiatric diagnoses reflected the usual range of patient material encountered at the clinic.

None of the differences between the scores of the two administrations approached significance. Both in pattern and amplitude pre- and posttests were almost identical.

The voluminous body of literature utilizing the MMPI to demonstrate differences between groups and, even more sensitively, changes in scores following psychotherapy suggests caution in cavalierly discarding these data and turning to the parsimonious challenge of the validity of the technique. In any case, these results furnish some food for thought.

Brief Report.
Received January 30, 1958.

Reference

1. Greenfield, N. S., & Fey, W. F. Factors influencing utilization of psychotherapeutic services in male college students. *J. clin. Psychol.*, 1956, 12, 276–279.

---

[1] An extended report of this study may be obtained without charge from Norman S. Greenfield, Department of Psychiatry, University Hospitals, Madison 6, Wisconsin, or for a fee from the American Documentation Institute. Order Document No. 5584, remitting $1.25 for microfilm or $1.25 for photocopies.

[2] Some of the data for this study were made available through the generous cooperation of Lewis E. Drake.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

The Diagnostic Validity of the Ames "Danger Signals"

Elise Elkins[1]
Illinois Institute for Juvenile Research

The application of the Rorschach to very young subjects has resulted in an accumulation of studies indicating that children normally give responses that are pathological in adults (10, 13). The publication of *Child Rorschach Responses* by Ames and her associates (1) represented a milestone in that developmental norms were presented systematically at yearly intervals for a carefully studied group of children ten years of age and under. Toward the end of this volume, Ames presented a list of "danger signals" which, even within the developmental context of a child's Rorschach, might still be considered indicative of pathology. The present study was designed to test whether these "danger signals" had sufficient diagnostic validity to differentiate between a clinic and a normal population. As part of a study of normal eight-, nine-, and ten-year-olds undertaken by the Institute for Juvenile Research, 54 Chicago school children described as well-adjusted by their teachers were given a complete diagnostic evaluation. For the purposes of this study, an effort was made to match each of these nonclinic cases with a child who had been referred to the Institute as being in need of child guidance services. The present report deals with a comparison of the Rorschach records of 40 nondisturbed children with the records of their matched controls.

Subjects

The *S*s were 40 normal children (27 boys and 13 girls) and their controls who were matched for sex, age, and IQ. Matching was in most cases within three months of age and five points of IQ. The matching for IQ was only approximate. All of the nondisturbed *S*s were given Form L of the Binet by the same examiner, while their matched controls were given a variety of other tests, both group and individual. In all likelihood, the clinic group is slightly more intelligent than the nondisturbed sample, since the clinic children were often matched on the basis of tests scores that are known to run lower than Binet ratings. This direction of a possible difference was considered safest so that there would be no question of the clinic children's looking "sick" only because they were merely less intelligent. Actually, both the clinic and the nondisturbed children taken together constitute a highly homogeneous sample of children of superior intelligence. In this respect, both groups are comparable to the normative sample used by Ames (1, p. 25).

Although no effort was made to match for socioeconomic status, the matching for IQ resulted in a clinic and a nondisturbed sample which were quite comparable in this regard. The socioeconomic level of the *S*s was evaluated by means of the scale of paternal occupations described by Goodenough (6, pp. 501–512). Three clinical psychologists classified the 80 *S*s. In 39 of the 80 cases, all three judges agreed on the classification. In an additional 31, one judge differed by one category from the other two. The *S* was then assigned to the category agreed upon by two out of three judges. In the nine cases with ratings spread over three categories, the *S* was assigned to the middle category. Both the clinic and the nondisturbed group are slightly below the Ames sample in regard to socioeconomic status: 93% of the Ames sample (1, p. 25) and 59% of the combined

---

[1] The author would like to express her gratitude to Charlotte Altman who tested the entire nondisturbed sample; to Eugene Levitt who gave helpful suggestions; and to Irene Goldblatt, Jeanne Hershenson, and Victor Lebow who served as scorers and raters.

clinic and normal groups of this study fall into Categories I through III on the scale of paternal occupations. The atypically high socioeconomic status of the Ames sample in comparison to the general population raises the possibility of bias in the Ames normative data. In fact, Fiedler and Stone (5, p. 279) found striking differences between their sample of undisturbed children of low socioeconomic status and the Ames sample of high socioeconomic status. However, in this study, any possible effects of bias in the Ames norms are controlled by the matched-pairs design.

Procedure

All 40 Rorschachs of the nondisturbed children were administered and scored by a psychologist with 15 years' experience and certification by the American Board of Examiners in Psychology. The 40 Rorschachs of the matched controls were administered and scored by 15 psychologists of varying levels of experience. Because of the extreme difficulty of obtaining enough clinic cases to match with a group of normals of such high intelligence, it was not possible to control for experience or identity of the examiner. Since arrangements could not be made to have all 80 Rorschachs retyped without identifying information, this study was limited to objective signs whose presence or absence could be evaluated merely by a check of the scoring sheet.

Table 1

Descriptive Information About Subjects Used in This Study

Variable	Nondisturbed Group $N = 40$	Clinic Group $N = 40$
Age		
Range	8-4—10-0	8-3—9-11
Mean	108.38 mos.	108.15 mos.
SD	5.50 mos.	5.40 mos.
Sex		
Male	27	27
Female	13	13
IQ		
Range	98–145	100–155
Mean	120.15	120.30
SD	12.52	14.06
Socioeconomic status		
I. Professional	6	8
II. Semiprofessional and managerial	6	5
III. Clerical, skilled trades, retail business	11	11
IV. Semiskilled, minor clerical	11	9
V. Slightly skilled	4	4
VI. Day laborers	2	3

Of the 13 "danger signals" listed by Ames and her associates, 4 were dropped because they involved subjective judgments. The remaining 9 were raised to a total of 13 because of the necessity of treating the parts of four complex signs as separate signs (e.g., Sign 10: *A*% over 60 and no human responses). The 13 signs used in this study are listed below with the numbering used by Ames (1, p. 283). Qualifications introduced by the author are given in parentheses.

1. Preponderance of tree or leaf responses (Preponderance defined as over 50% of *R*)
2. *F*% exceeds 75
3. More than one pure color response
5A. More than one response involving differentiated shading
5B. More than one response involving diffuse, dysphoric shading, i.e., *Clob*
6. More than one inanimate movement response
8. W% less than 40
9A. *Dd*% over 15. *Dd*% equals *Dd, S,* and *Do* responses over total *R*
9B. More than one *X* response
9C. More than one *S* response
10A. *A*% over 60
10B. No human response (i.e., no *H: Hd* not counted)
11. *R* less than 10

The evaluation of the 13 signs was done by the author. A slight element of subjective judgment could not be avoided in that some rescoring had to be done before even the objective Ames signs could be checked. Of the 80 protocols used in the study, 70 were scored by the Beck system (2), while the Ames system is closer to the Klopfer method (9). Fortunately, it has been common practice at the Institute to make marginal notations of *m* and *FM*. However, the elimination of separate *D* scorings for details covered in a *W* and the division of shading scores into *Clob* (diffuse, dysphoric) and *F*(*C*) (differentiated) also had to be done. The author prepared a detailed instruction sheet indicating the differences between the Ames and the usual clinic scoring in regard to those variables that would be involved in an evaluation of the Ames danger signals. The scoring record for each of the 80 protocols was then transferred to a new scoring sheet, with each response being scored only for the relevant variables which are as follows:

Area (*W*, *D*, *Dd*, *Do* response, *S*)
Major determinant (*M*, *m*, *FM*, *C*, either *FC* or *CF*, differentiated shading, diffuse shading with dysphoric element, *F*)
Content (*H*, *A* or *Ad*, tree or leaf)

No differentiations other than those listed were made. For example, no effort was made to differentiate between *FC* and *CF*. The Ames signs were then checked by counting the various types of responses on these scoring sheets. Tabulation of the Ames signs was, of course, purely clerical once the scoring sheets had been made out.

As for the possibility of the introduction of bias in the rescoring, the author can state only that there seems to be satisfactory evidence in the literature that mere scoring of Rorschach responses has sufficient interexaminer reliability to warrant being regarded as an objective procedure (7, 8). However, as a reliability check, one-fourth of the protocols (10 normals and 10 clinic cases) were scored by a colleague as well as by the author. The percentage of agreement on the scoring of location varied from 71 to 93 on the 20 records, with a mean of 81 and an average deviation of 5.25 percentage points. The percentage of agreement on the scoring of major determinant varied from 65 to 100 on the 20 records, with a mean of 84 and an average deviation of 7.05 percentage points. The percentage of agreement on the scoring of content varied from 77 to 100 on the 20 records, with a mean of 91 and an average deviation of 6.10 percentage points. There was perfect agreement on content scoring for a third of the records.

Table 2

Total Scores on Ames "Danger Signals" and Davidson Adjustment Signs in Matched Pairs of Clinic and Nondisturbed Subjects

	Nondisturbed Group	Clinic Group	*CR*
Ames "Danger Signals"			
Mean	3.00	2.20	2.25*
SD	1.55	1.75	
Davidson Adjustment Signs			
Mean	7.10	6.58	.57
SD	2.27	2.84	
Prorated mean	9.26	8.60	

* Significant at .05 level.

Results

Total Scores

The total scores on the Ames list of danger signals were compared for the clinic and the nondisturbed groups by means of a variation of McNemar's Formula 26 for the evaluation of the difference between two correlated means (12, p. 74). The results are presented in Table 2. The clinic group had significantly fewer danger or maladjustment signs than the nondisturbed group. Before investigating the direction of the discrimination done by each individual Ames sign, the author compared the clinic and the nondisturbed groups by means of 13 of the 17 Davidson adjustment signs (4). The two signs involving color shock and shading shock respectively were omitted because they involve subjective judgment. The two signs involving *FM* and *F*% respectively were omitted because the records were scored by the Beck system. The Davidson signs were evaluated on the basis of the scoring done by the original examiners without any rescoring by the author. The results of this comparison are also presented in Table 2. In contrast to the Ames signs, the Davidson signs show an absolute difference in the expected direction; however, the difference is so small that there is a high probability that it could be a chance deviation from a true difference of zero ($p > 60$). Although all statistical tests were done on the basis of the 13 signs actually counted for both the nondisturbed and the clinic groups, the mean based on 13 signs was prorated to obtain the estimated mean had all 17 signs been used. The prorated means are also shown in Table 2. An estimate of the accuracy of the prorating is provided by some additional data available for the nondisturbed sample only. The original examiner had evaluated the group on the basis of all 17 signs and obtained a mean of 9.3. The prorated mean obtained by extrapolating from the mean of 7.1 for 13 signs is also 9.3. The prorated means of 9.3 and 8.6 for the nondisturbed and the clinic groups respectively are consistent with the results of previous studies which show slight differences between normal and disturbed groups with considerable overlapping of the distributions of scores. In the studies reviewed by Davidson, means of 9.05 to 11.34 were reported for better-adjusted groups and means of 6.85 to 8.17 for less well-adjusted groups (4).

Individual Signs

The Davidson signs (4) did not differentiate the nondisturbed from the clinic patients in the present study, and the Ames danger signals discriminated in the "wrong" direction when evaluated in terms of total scores. Because of the possibility that individual Ames signs discriminated in different directions, the author carried out a detailed analysis, sign by sign. The results of this analysis are presented in Table 3.

Table 3

Incidence of Ames "Danger Signals" in Matched Pairs of Clinic and Nondisturbed Subjects

Danger signal	Both of pair	Neither of pair	Only clinic	Only Nondisturbed	Chi square
1A Preponderance of tree and leaf R	0	40	0	0	0
2 $F\% > 75$	1	26	4	9	1.23
3 Pure $C > 1$	0	38	2	0	.50
5A Differentiated shading > 1	3	21	6	10	.56
5B Diffuse, dysphoric shading > 1	0	36	3	1	.25
6 $m > 1$	0	36	4	0	2.25
8 $W\% < 40$	17	3	6	14	2.45
9A $Dd\% > 15$	12	12	4	12	3.06
9B X responses > 1	0	33	5	2	.57
9C $S > 1$	2	33	2	3	0
10A $A\% > 60$	2	16	5	17	5.50*
10B No H	4	24	4	8	.75
11 $R < 10$	0	39	0	1	0

* Significant at .05 level.

In the case of each sign, the number of instances of occurrence in the nondisturbed group and lack in the clinic cases were compared with the number of instances of occurrence in the clinic group and lack in the nondisturbed cases by means of McNemar's formula for chi square for dependent samples (12, pp. 204–206). The use of McNemar's formula permits one to evaluate only those two cells of the fourfold table which represent the difference critical for one's hypothesis. In this case, only the third and fourth columns of Table 3 were compared. The Yates correction for continuity was used in all computations.

The data in Table 3 suggest first of all that most of the Ames danger signals are not useful for differentiating between clinic and nondisturbed samples of eight- to ten-year-olds because nearly half of the signs are either characteristically present in a sizable proportion of both normal and clinic cases or else are almost never found in the records of either. Thus in 42% of the matched pairs, both the clinic and the nondisturbed member have a $W\%$ below 40. In 100% of the matched pairs, neither member has a preponderance of tree and leaf responses. In 90% of the matched pairs, neither member has more than one response involving diffuse, dysphoric shading. The only "danger sign" which has differentiating power that reaches statistical significance is Sign 10A ($A\%$ over 60). This sign discriminates in an unexpected direction: Of 22 matched pairs whose members differ on this sign, three times as many nondisturbed as clinic cases have the high $A\%$. The most promising signs which discriminate in the expected direction are Signs 3 (pure $C > 1$), 5B (more than one instance of diffuse, dysphoric shading) and 6 ($m > 1$). These signs, like contamination for schizophrenia, are apparently diagnostic when they occur, but have such limited incidence as to be of little use for routine screening purposes.

Ranges on Determinants Involved in the Ames Danger Signals

The Ames signs, in contrast to the Davidson signs, make less allowance for ranges of scores with an optimum somewhere in the middle. Therefore, at the risk of inflating

Table 4

Comparison of Matched Pairs of Clinic and Nondisturbed Subjects in Regard to Scores Below the 25th Percentile of the Ames Norms

Determinant and level necessary to fall below 25th percentile	Both of pair below	Neither of pair below	Only clinic below	Only nondisturbed below	Chi square
$F\%$ below 52	4	20	8	8	0
No pure C	31	0	4	5	0
Differentiated shading zero	10	15	5	10	1.07
Diffuse, dysphoric shading zero	17	1	3	19	10.23**
m of zero	20	4	3	13	5.06*
$W\% < 28$	11	8	7	14	1.71
$Dd\%$ zero	1	31	6	2	1.12
$A\% < 40$	3	24	6	7	0
$R < 13$	0	33	2	5	.57

* Significant at .05 level.
** Significant at .01 level.

probabilities, one further analysis of the data was performed. The matched pairs of clinic and nondisturbed cases were compared in regard to the determinants underlying the danger signals in two ways: (*a*) frequency at or below the 25th percentile [2] of the Ames norms and (*b*) frequency above the 75th percentile of the Ames norms (1, pp. 54–56). These findings are presented in Tables 4 and 5.

Table 4 indicates that in the case of those matched pairs whose members differ in regard to being below the 25th Ames percentile, there are significantly more pairs whose nondisturbed member is below the 25th percentile for number of *m* and for number of responses involving diffuse, dysphoric shading. (As before, only Columns 3 and 4 were compared by means of McNemar's formula for chi square for dependent samples.) Table 5 indicates that in the case of those matched pairs whose members differ in regard to being above the 75th Ames percentile, there are significantly more pairs whose nondisturbed member is above the 75th percentile for *A%* and significantly more pairs whose clinic member is above the 75th percentile for number of responses involving diffuse, dysphoric shading.

A comparison of Tables 4 and 5 with Table 3 lends no support to the author's suspicion that the Ames danger signals would be improved by the introduction of middle ranges of scores on various determinants. Three of the four significant differences in the latter two tables result from a change in the cutting score for variables which appear to have a linear relationship to good adjustment. Thus, a complete lack of *m* and of dysphoric shading, rather than the presence of a moderate amount, differentiates the clinic from the nondisturbed sample. In the case of the fourth significant difference on *A%*, the data do not reveal whether there is any point at which the relationship between health and *A%* ceases to be linear in regard to the groups used in this study. The highest *A%* of 93 was found in the nondisturbed group, in addition to the fact that more of the undisturbed group exceeded the Ames 75th percentile (*A%* of 58).

Discussion of Results

The Ames danger signals offer no advantages over the other sign lists developed without special attention to the younger age groups and the developmental sequences particular to them. Neither are the Ames danger

Table 5

Comparison of Matched Pairs of Clinic and Nondisturbed Subjects in Regard to Scores Above the 75th Percentile of the Ames Norms

Determinant and level necessary to exceed 75th percentile	Both of pair above	Neither of pair above	Only clinic above	Only nondisturbed above	Chi square
$F\%$ above 78	2	26	3	9	2.08
Pure $C > 1$	0	38	2	0	.50
Differentiated shading over 1	3	21	6	10	.56
Diffuse, dysphoric shading > zero	1	18	18	3	9.33**
$m > 1$	0	37	3	0	1.33
$W\% > 52$	1	27	6	6	0
$Dd\% > 16$	11	12	4	13	3.76
$A\% > 58$	4	15	4	17	6.86**
$R > 23$	6	16	11	7	.50

** Significant at .01 level.

---

[2] In the case of those signs with zero as the score for the 25th percentile, comparisons had to be made in regard to comparative frequencies at the 25th percentile since, of course, there could be no frequencies below zero.

signals any worse for screening purposes. The fact that several of the Ames signs discriminate in a direction opposite to that expected has been found to be true of comparable sign lists such as the Davidson adjustment signs. Brockway, Gleser, and Ulett (3), for example, found that 8 of the 17 Davidson signs discriminated in a reverse direction when used on a population of adult males.

One of the studies most comparable to the one described in this paper was carried out at the same clinic by another investigator. Levitt (11) compared 39 clinic cases, including 4 used by the present author, with the normative sample of Thetford, Molish, and Beck in regard to alleged Rorschach anxiety indices. He found that only *Sum Y* discriminated in the expected direction at better than the .05 level of confidence. The discriminating power of *Y* was improved by separating the differentiated shading responses (*FY*) from those with less of a form element (*YF*). Similarly, in the study by the present author, the best discrimination was obtained by the use of *Clob* responses only, the latter being roughly comparable to the *YF* responses of Levitt's study. However, the sum of all shading responses had some discriminating power. When the 40 matched pairs of clinic and nondisturbed subjects were compared in regard to total number of shading responses, significantly more pairs were found with the normal subject at or below the Ames 25th percentile[3] of zero shading responses (chi square 5.26 and $p < .05$). When a similar comparison was made in regard to scores above the Ames 75th percentile, there was a statistically insignificant trend for more clinic than normal members of matched pairs to be above the Ames 75th percentile for total number of shading responses. That is, complete lack of shading responses is significantly associated with membership in the nondisturbed group, while having two or more shading responses is less discriminating. Once shading responses are present, it is necessary to know the type before meaningful discriminations can be made between groups.

In general, those signs and determinants which discriminate in any way between the clinic and the disturbed group seem to have to do with tension (*m*, shading responses) and the stereotypy-conformity dimension of behavior (*A%*). Both the clinic and the nondisturbed samples are children of latency age. The nondisturbed sample appears to be distinguished mainly by a consolidation of defenses. Apparently, deviation in the direction of more control than the normative Ames sample is not pathological at this age, whereas deviation in the direction of less control identifies the patient population. Although in the adult, magnitude of deviation may be of major importance, it is possible that in the child the direction of the deviation in terms of the child's developmental level is equally important. The nondisturbed group in this study is exaggeratedly like the conforming, controlled model child that is the societal norm for the latency period. The fact that the nondisturbed sample was chosen by school teachers undoubtedly increased the probability that a particular variety of the societal ideal of normal latency adjustment would constitute the sample. Perhaps in the expansive period of adolescence the deviations in the direction of overcontrol are the more pathognomonic than similar quantitative deviations in the direction of impulsivity.

Summary

Forty children described as well-adjusted by their teachers were matched for sex, age, and IQ with children referred for child guidance services to a large community clinic. Each group contained 13 girls and 27 boys. The average age of both groups was a little over nine years, and the average IQ of both groups was a little over 120. The matched pairs were compared by means of the Rorschach in the following ways: (*a*) total number of objectively scored Ames danger signals out of a list of 13, (*b*) total number of objectively scored Davidson adjustment signs out of a list of 13, (*c*) presence or absence of individual Ames danger signals, (*d*) compara-

---

[3] The Ames percentile for total number of shading responses was estimated by adding the number of differentiated and diffuse shading responses at the given percentile. Thus the 25th percentile for all shading responses was estimated as zero, since zero is the 25th percentile for both *F*(*C*) and *Clob*. The 75th percentile was estimated as 1 since the 75th percentile for *F*(*C*) is zero and for *Clob* is 1.

tive frequencies above the Ames 75th percentile on determinants underlying the Ames danger signals, and (*e*) comparative frequencies below the Ames 25th percentile for determinants underlying the Ames danger signals.

The nondisturbed group was found to have a significantly higher total of Ames danger signals. The difference between the two groups on the Davidson adjustment signs did not reach statistical significance, but was in the expected direction. When the individual Ames signs were evaluated separately, only *A%* above 60 was found to discriminate at a statistically significant level. The difference was in the reverse of the expected direction, i.e., the nondisturbed group had more members with the high *A%*. When comparative frequencies above and below the Ames 75th and 25th percentiles respectively were evaluated, only the following differences between the clinic and the nondisturbed group were found to be statistically significant ($p < .05$): fewer clinic children with no *m*, fewer clinic children with *A%* over 58, more clinic children with responses involving diffuse, dysphoric shading, and more clinic children having two or more of the latter type of response.

The author concluded that the Ames danger signals do not have sufficient diagnostic validity to make them useful for differentiating between maladjusted and nondisturbed samples of eight- to ten-year-olds. Nearly half of the signs are either characteristically present in a sizable proportion of both normal and clinic cases or else are almost never found in the records of either. Several of the "danger" signs actually tend to be found *more* frequently in normal records. However, the developmental frame of reference underlying the signs was still seen as useful in interpreting the findings of the study.

Received July 22, 1957.

References

1. Ames, Louise B., Learned, Janet, Metraux, Ruth W., & Walker, R. N. *Child Rorschach responses: Developmental trends from two to ten years.* New York: Hoeber-Harper, 1952.
2. Beck, S. J. *Rorschach's test.* Vol. I. *Basic processes.* New York: Grune & Stratton, 1949.
3. Brockway, Ann L., Gleser, Goldine, & Ulett, G. A. Rorschach concepts of normality. *J. consult. Psychol.,* 1954, **18**, 259–265.
4. Davidson, Helen H. A measure of adjustment obtained from the Rorschach protocol. *J. proj. Tech.,* 1950, **14**, 31–38.
5. Fiedler, M. F., & Stone, L. J. Rorschachs of selected groups of children and published norms: II. The effect of socio-economic status on Rorschach performance. *J. proj. Tech.,* 1956, **20**, 276–280.
6. Goodenough, Florence L., & Anderson, J. E. *Experimental child study.* New York: Century, 1931.
7. Hertz, Marguerite. Scoring the Rorschach inkblot test. *J. genet. Psychol.,* 1938, **52**, 15–64.
8. Hertz, Marguerite. On the standardization of the Rorschach method. *Rorschach Res. Exch.,* 1939, **3**, 120–133.
9. Klopfer, B., & Kelley, D. *The Rorschach technique.* Yonkers, N. Y.: World Book, 1946.
10. Klopfer, F., & Margulies, Helen. Rorschach reactions in early childhood. *Rorschach Res. Exch.,* 1941, **5**, 1–23.
11. Levitt, E. E. Alleged Rorschach anxiety indices in children. *J. proj. Tech.,* in press.
12. McNemar, Q. *Psychological statistics.* New York: John Wiley, 1949.
13. Paulsen, Alma. Rorschachs of school beginners. *Rorschach Res. Exch.,* 1941, **5**, 24–29.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

An Investigation of Determinants Inducing Individuals to Seek Personal Counseling[1,2]

James S. Terwilliger and Fred E. Fiedler
University of Illinois

This study attempted to identify some of the determinants which lead individuals to seek therapeutic help. Very few situations permit an adequate study of this problem. Generally the counselor only sees people who feel the need for psychotherapy, while many others, equally maladjusted by various criteria, choose, for various reasons, not to consult a therapist. It is, therefore, of considerable theoretical and practical importance to determine how those who seek counseling differ in important respects from those who do not.

Such a study was possible at the University of Illinois where the Student Counseling Service deals with students' educational, vocational, and personality problems. Students are urged to consult counselors if they, at any time, feel the need for professional help. Counseling services are provided without charge.

Sample. The subjects (*S*s) were male University freshmen who lived in temporary barracks which normally housed 13 to 14 men. *S*s are assigned to the barracks in the order in which their application is received.

On the basis of counseling service records, we identified a "counseling bureau group" (CBG) consisting of *S*s who had visited the Counseling Bureau two or more times for personality problems ($N = 10$). Those who had never consulted the bureau were placed into the control group ($N = 87$).

Tests. A number of personality and adjustment measures were obtained from each *S*. These included an abbreviated Taylor Manifest Anxiety Scale, a self-esteem score, a self-ideal self discrepancy (or self-satisfaction) score. Also obtained were sociometric status indices to provide a measure of the individual's acceptability to his peers, and "esteem by others" scores from *S*'s fellow students who lived in the same barracks. This latter score indicates *S*'s adjustment as evaluated by others. In addition, we also computed a Grade Point Average, adjusted for *S*'s performance in high school.

Results. Taylor anxiety scores, self-satisfaction, and self-esteem scores of *S*s who consulted therapists differed significantly in the maladjusted direction from those who did not consult a therapist. However, differences in other scores were not significant.

Discussion. As is to be expected, anxious and less self-satisfied individuals are more likely to seek therapeutic help than are those who do not express such feelings. In contrast, however, these *S*s do not differ from the control group on objective or outside indices. *S*s who consulted therapists were as well liked as others, they were esteemed as highly, and they tended to have slightly better grades. These findings suggest that the client is brought into therapy solely by subjective feelings of discomfort with himself rather than by "reality" factors.

Brief Report.
Received February 20, 1958.

---

[1] This study was conducted under Contract DA-49-007-MD-569 between the Office of the Surgeon General, Department of the Army, and the University of Illinois.

[2] An extended report of this study may be obtained without charge from Fred E. Fiedler, 427 Gregory Hall, University of Illinois, Urbana, Illinois, or for a fee from the American Documentation Institute. Order Document No. 5582, remitting $1.25 for microfilm or $1.25 for photocopies.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

Vocabulary as a Basis for Estimating Intelligence from the Rorschach[1]

Thomas R. Trier
University of California, Berkeley

An estimate of intelligence is usually included in the formulations made from the Rorschach. It can be particularly helpful when time limitations or other considerations render it impractical or impossible to administer an intelligence test. Beck (2) and Klopfer and Kelley (3) are among those who have tried to make more explicit the variables which should be considered in estimating intellectual level. Beck states that intelligence is related to Z, $A\%$, $F+\%$, Ap, and number and quality of M responses. Klopfer and Kelley suggest that an estimate of intelligence should be based on number and quality of W, M, and O responses, form accuracy level, variety of content, and succession of responses. Research seeking to determine whether the postulated relationships between Rorschach variables and IQ do in fact exist has been rather unsuccessful. Usually, when low, significant correlations between Rorschach factors and IQ have been found, they have failed to hold up under cross-validation. The results reported by Armitage *et al.* (1) are typical in this respect, but are particularly worthy of mention because of the large number of cases employed. The authors computed the correlations between 19 Rorschach variables and Wechsler-Bellevue IQ for 503 patients. Low correlations, significant at the .01 level, were obtained for 13 of the variables. A multiple regression equation, based on the six variables showing the highest correlation with IQ, was computed. When this equation was applied to a new sample of 207 patients, it failed to improve the accuracy of estimates beyond chance. In contrast, judges were able to make fairly accurate estimates of intelligence when exposed to the total protocols. The authors suggest the possibility that the judges were responding to the quality of the vocabulary and the quality of the perceptual organization in making their judgments.

The present study is an attempt to determine whether the vocabulary in Rorschach protocols is a variable that can be profitably utilized in estimating intelligence. Wechsler's statement, ". . . the size of a man's vocabulary . . . is an excellent measure of his general intelligence" (6, p. 98), has been generally accepted. Vocabulary subtests are included in many intelligence tests. In both the Stanford-Binet (4) and Wechsler-Bellevue (6), the vocabulary subtest correlates higher with total IQ than do any of the other subtests. If the extensiveness of a person's vocabulary can be judged from the Rorschach with some accuracy, this measure might be useful in estimating his intelligence.

Procedure

Three groups, each consisting of four judges, were asked to make numerical estimates of Wechsler-Bellevue IQ. Judges in Group I were given 16 Rorschach protocols. They were asked to list the seven "most sophisticated" words from each protocol and to base their IQ estimates on these word lists. Judges in Group II were each given one of the four sheets of word lists constructed by the judges of Group I. They were informed how the lists were derived and were asked to make IQ estimates using only the word lists. They differed from Group I in having no access to the total protocol. Group III judges were handed the sixteen protocols and asked to estimate intelligence from them. In contrast to the others, Group III judges were not given specific instructions on what to base their judgments.

---

[1] This research was performed while the writer was a psychology trainee at the Veterans Administration Hospital, San Francisco, California. He wishes to express his appreciation to David A. Rodgers for his valuable assistance in the design and analysis of the experiment and in the preparation of the article.

In addition, an attempt was made to measure the extensiveness of vocabulary in a more objective fashion. The four word lists constructed by Group I judges for each protocol were pooled. While these word lists overlapped considerably, pooling revealed that from 10 to 20 different words had been selected as "most sophisticated" for each protocol by the four judges, with a mean of 14.7. These words were looked up in Thorndike and Lorge's *The Teacher's Word Book of 30,000 Words* (5) to determine how frequently each word appears in everyday usage. A number that indicated how often each word appears among a million written words was obtained. The seven most infrequently appearing words were then selected for each protocol and the frequencies of these seven words were totalled. Sixteen sums were thus obtained, one for each protocol. Rank-order estimates were made using the hypothesis that there would be an inverse relationship between the size of the sum and IQ.

The criterion measures of intelligence were based on full scale IQs on the Wechsler-Bellevue scale. Four Rorschach records were obtained from patients in each of the following IQ ranges: 120 and above, 110–119, 90–109, and 89 and below. Thus IQs were roughly rectangularly distributed, permitting comparisons of the accuracy of judgments based on the total protocols (Group III) with the results reported by Armitage *et al.* (1), who used a similar distribution. The protocols of male veterans who were referred for psychological evaluation during hospitalization were selected from the files at San Francisco Veterans Administration Hospital. The tests had been administered by eight different examiners. Patients suspected of having organic brain pathology were excluded from the sample. Only those cases whose Rorschach protocols were scored by the Klopfer and Kelley system (3) were used. The age of the patients in the sample ranged from 19 to 67, with a mean of 38.4. The mean number of years of education was approximately ten. From a diagnostic standpoint, the group was heterogeneous and included psychoneurotic, psychotic, personality disorder, and psychosomatic cases.

The 12 clinicians used as judges all had extensive training and experience in the use of the Rorschach and Wechsler-Bellevue. Most of them were graduate students nearing the completion of their Ph.D.s in clinical psychology.

Table 1

Rho Correlations Between the IQ Estimates of Judges and Wechsler-Bellevue IQ

Judge	Group I	Group II	Group III
1	.79**	.85**	.79**
2	.76**	.81**	.67**
3	.74**	.66**	.64**
4	.68**	.63**	.61*

* Significant at or below the .05 level.
** Significant at or below the .01 level.

Results

Correlation coefficients were used as indicators of the accuracy of judgments. In view of the rectangularity of the distribution, rank-order coefficients were employed. Table 1 indicates the correlations and the corresponding significance levels which were obtained by the judges. Results for Groups I and II indicate that estimates of intelligence can be made from Rorschach vocabulary. Since the accuracy of judges in Groups I and II is approximately the same, it does not appear to make any difference whether judges making their estimates from word lists have access to the protocols or not. It is to be noted that the judgments based on total protocol are no more accurate than those based solely on the vocabulary. In addition, the estimates based on the Thorndike and Lorge word frequency count resulted in a rho of .77 ($P < .01$), which is roughly comparable to the accuracy achieved by the judges. The results for Group III are in agreement with those of Armitage *et al.* (1), who report a median correlation of .69 for estimates based on evaluation of the total protocols.

Discussion

In comparison to the conventional Rorschach scoring variables whose contributions toward estimating intelligence have been quite limited, vocabulary is a variable that improves accuracy of judgments considerably. To rely solely on vocabulary in estimating IQ would simplify the task and might well be the most practical approach at present. The results suggest that the clinician is not indispensable for making the estimates. The Rorschachs could be submitted to a clerk who is instructed to select the rarest words from each protocol and to look them up in a word frequency table. After finding the seven most uncommon words in each, he totals their frequencies and enters this total into an empirically derived equation or table, which yields a numerical estimate of IQ.

The present investigation points the way toward areas of possible further research:

1. It is quite likely that vocabulary could also be evaluated from samples of verbal behavior other than the Rorschach. Thus estimates of intelligence might be made from the Thematic Apperception Test or from interviews, possibilities first suggested by Armitage *et al.* (1).

2. Since the Thorndike and Lorge word list is somewhat outdated and contains only 30,000 words, it is to be expected that if a more current and extensive word list could be utilized, it would lead to somewhat more accurate estimates of IQ than those reported here.

3. Since judges in Group III were not limited to specific cues in making their estimates of IQ, it is impossible to determine to what extent their accuracy of judgment was due to reliance on vocabulary. It would be informative to determine whether comparable accuracy would be achieved when vocabulary is eliminated as a cue. This could be accomplished by translating the protocols into basic English before giving them to the judges.

Summary

Three groups, each consisting of four judges, made estimates of Wechsler-Bellevue IQ from 16 Rorschach protocols. One group used word lists which they constructed from the protocols, another used word lists derived by others from the protocols, and the third used the protocols without instructions to focus on vocabulary. In addition, estimates were made in a more objective fashion using the Thorndike-Lorge word frequency count. The three groups and the objective approach achieved fairly comparable accuracy. Results suggest that IQ can be estimated as well, using only vocabulary, as by any other approach currently in use.

Received August 15, 1957.

References

1. Armitage, S. G., Greenberg, P. D., Pearl, D., & Daston, P. G. Predicting intelligence from the Rorschach. *J. consult. Psychol.*, 1955, **19**, 321–329.
2. Beck, S. J. *Rorschach's test*. Vol. II. *A variety of personality pictures*. New York: Grune & Stratton, 1949.
3. Klopfer, B., & Kelley, D. M. *The Rorschach technique*. Yonkers, N. Y.: World Book, 1942.
4. Terman, L. M., & Merrill, M. A. *Measuring intelligence*. Boston: Houghton Mifflin, 1937.
5. Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words*. New York: Teachers College, Columbia Univer., 1944.
6. Wechsler, D. *The measurement of adult intelligence*. (3rd ed.) Baltimore: Williams & Wilkins, 1944.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

Extraversion, Neuroticism, Radicalism, and Tendermindedness[1]

A. W. Bendig
University of Pittsburgh

Eysenck (2) has stated that the personality trait of Extraversion is negatively related to the social attitude dimension of Tendermindedness, but is not related to the attitude trait of Radicalism. Neuroticism is not related to either attitude dimension. However, Cattell states that Extraversion is negatively related to Radicalism (1, pp. 317–318) and also implies a relationship between Neuroticism and Radicalism (1, p. 209), but the direction of the correlation is left unclear. The present research was designed to clarify some of the interrelationships among these four variables.

The Maudsley Personality Inventory, containing scales of Extraversion and Neuroticism, was administered to 208 undergraduate *Ss* (117 men and 91 women) in five sections of introductory psychology. The scales of Radicalism and Tendermindedness developed by Melvin (2, pp. 276–280) were also administered to the same *Ss*. Comparisons of the mean raw scores of the two sex groups showed the female *Ss* to be more "tenderminded" than the male *Ss*, but the differences on the other three scales were not statistically significant.

Product-moment correlation coefficients among the four scales were separately computed within each sex group and the pairs of correlations tested for significant sex differences by the r-to-z method. The correlations were averaged when these tests did not approach significance. The average correlation between Extraversion and Neuroticism ($r = -.12$) was significant at the .10 level, but the correlations of Extraversion with Radicalism and Tendermindedness were not significant ($r = -.04$ and $-.01$). The average correlation between Neuroticism and Radicalism ($r = .22$) was significant at the .01 level and similar in both sex groups ($r = .18$ and $.26$). The correlations of Tendermindedness with Neuroticism and Radicalism were significant for the male *Ss* ($r = -.28$ and $-.19$), but not for the female *Ss* ($r = -.03$ and $.08$), and the sex differences in these correlations were significant at the .10 level ($z = 1.79$ and 1.95).

These results indicate that Eysenck's linking of Extraversion with Tendermindedness and Cattell's association of Extraversion and Radicalism were not confirmed by the data. However, Cattell's hypothesized relation of Neuroticism and Radicalism did appear within both sex groups of *Ss*. The sex differences among the correlations of Tendermindedness with Radicalism and Neuroticism may represent either a sampling fluctuation or a different organization of these three traits in men and women *Ss*.

Brief Report.
Received March 7, 1958.

References

1. Cattell, R. B. *Personality and motivation structure and measurement*. Yonkers, N. Y.: World Book, 1957.
2. Eysenck, H. J. *The psychology of politics*. London: Routledge & Kegan Paul, 1954.

---

[1] An extended report of this study may be obtained without charge from A. W. Bendig, Dept. of Psychology, University of Pittsburgh, Pittsburgh 13, Pa., or for a fee from the American Documentation Institute. Order Document No. 5578, remitting $1.25 for microfilm or $1.25 for photocopies.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

Handwriting Conventions as Determinants of Human Figure Drawings[1]

Wayne Dennis
Brooklyn College

Projective theorists have suggested interpretations for many aspects of human figure drawings, among them the position of the drawing upon the page. Thus, Karen Machover (2) proposed that the location of the drawing toward the right side of the page indicates that the subject is environment-oriented, while a left-side location is characteristic of a self-oriented person. A drawing high on the page is related to optimism, a low-positioned drawing suggests depression. Levy (1) proposed that drawings on the left of the page indicate self-consciousness or introversion, toward the right, rebelliousness. For Levy, the use of the upper half of the page signifies striving toward high standards of achievement whereas persons who place drawings toward the lower edge of the sheet are likely to be stable, firmly rooted, or, in some cases, depressed. That these interpretations are based upon resemblances in meaning between words used to describe space (high, low, etc.) and words used to describe personality (elated, depressed, etc.) seems obvious. Whether such linguistic resemblances provide a valid basis for hypothesizing relationships between drawings and personality has not as yet been adequately determined.

Whatever may be the final decision in regard to the relationships between personality and graphic expression, it seems likely that not all of the variance present in drawings will be accounted for by individual differences in temperament and personality. There may be determinants of drawing position which are cultural and which have little if any relationship to personality. Culture determines the position on a page at which a person begins to write. If there is a habit-transfer between writing and drawing, the position of drawings will be affected by handwriting habits. This study examines the drawing location of two classes of writers; namely, those who begin to write at the top left of the page and those who begin to write at the top right of the page.

Method

Subjects

The writers who start at the top left of the page are represented by two subgroups, those writing English and those writing Armenian. In nationality, these *S*s are Americans and Lebanese, respectively. Those starting at the top right of the page are Arabic writers. Their nationalities are Lebanese and Egyptian. All are children in elementary schools between ages five and ten.

The American *S*s come from three widely separated localities: (*a*) a small town in Maryland, (*b*) a small town in Kansas, and (*c*) a suburban town near New York City. The Armenian *S*s attended two schools in Beirut, Lebanon. The *S*s comprised either complete samples or random samples from their respective schools. The total number of *S*s in the above groups was 1,075.

[1] We wish to express our appreciation to the Lebanese, Egyptian, and American schools who provided data for this study. The Egyptian and Lebanese data were gathered in 1956–1957 when the author was a visiting professor at the American University of Beirut, and the gathering and analysis of these data were supported by a grant from the Rockefeller Brothers Fund. Special thanks are due to Habib Kurani and E. Terry Prothro of the American University of Beirut for their cooperation and encouragement and to Evelyn Raskin of Brooklyn College for consultation and suggestions.

Some of the Arabic-writing *S*s were attending a public school in Port Said, Egypt. All others were attending public or private schools in Lebanon. These two groups totalled 651 *S*s. These *S*s, like the Armenian and American groups were, so far as could be determined, representative of their schools.

Testing and Scoring Procedures

The drawing test employed was the Goodenough draw-a-man test. Goodenough's instructions were followed. All drawings were made in pencil on 8½″ by 11″ white paper.

In scoring each drawing for the purposes of the present study, it was indicated, first, whether the topmost part was nearer or the bottom-most part was nearer their respective edges of the paper. Second, it was noted whether the left or right side of the drawing (left and right being designated from the observer's point of view) was nearer the lateral edges of the page. In judging the vertical location, the most extreme part of the drawing, such as the top of the hat or a projecting feather, was used. However, right and left judgments were based upon the most lateral parts of the *body*, excluding the extremities. In those drawings which lacked a body, judgments were based upon the head. When there was any doubt as to the judgments, measurements were made. When two measurements were identical, which seldom occurred, scores were assigned alternately to the two categories.

Results

In terms of the scoring method just described, a drawing must receive one of four scores: top left, top right, bottom left and bottom right. Top left means that the drawing was nearer the top of the paper than to the bottom, and nearer the left edge than the right edge. The other terms have corresponding meanings.

Table 1 shows the percentage of drawings falling into each category for each of the groups.

Table 1 shows that in the Armenian-American group, 48% of the drawings have a top-left position, whereas in the Arab group, 48% fall in the top-right group.

Table 1

Position of Figure Drawings on a Page

Location	Armenian-American	Arab
Top left	48%	24%
Top right	16%	48%
Bottom left	27%	14%
Bottom right	9%	14%
N	1,075	651

Discussion

Obviously, these findings are in accord with the habit-transfer hypothesis suggested earlier. Also in accord with this interpretation is the finding that 64% of the Armenian and American drawings and 72% of the Arab drawings are nearer the top of the page than they are to the bottom.

However, while *all* Armenians and Americans begin to write in the top-left portion of the page and *all* Arabic-writing *S*s start their handwriting in the top-right quadrant, not all members of these groups place their drawings in the corresponding positions. Handwriting customs appear to have some influence in producing differences between certain groups, but they cannot, of course, explain the within-group variability. While the data here presented are consonant with the handwriting hypothesis, they are not entirely conclusive. The two groups compared no doubt differ in other respects as well as in regard to handwriting. It is unlikely that one can find two groups differing in handwriting but identical in all other respects. However, if one were to find that many groups differing in handwriting also differ in a consistent way in location of drawings, our hypothesis would seem to be firmly supported. Until a wider sampling of groups is obtained, the data here presented should serve as a caution in regard to interpreting cross-cultural differences in drawing locations as indicative of differences in personality. Such differences may be the result of habit patterns having little relationship to emotional or social traits.

Summary

Evidence has been presented which is consonant with the hypothesis that the location of drawings depends in part upon whether subjects in writing their primary language begin at the top left or top right of the page. Further comparisons of other language groups are needed to test the hypothesis, but, meanwhile, caution in the interpretation of cross-cultural differences in location of drawings is indicated.

Received July 23, 1957.

References

1. Levy, S. Figure drawing as a projective test. In L. E. Abt & L. Bellack (Eds.), *Projective psychology*. New York: Knopf, 1950. Pp. 257–297.
2. Machover, K. *Personality projection in the drawing of the human figure*. Springfield, Ill.: Charles C Thomas, 1948.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

The Relationship Between Edwards' and McClelland's Measures of Achievement Motivation[1]

Levon H. Melikian
American University of Beirut

Independent measures of achievement motivation have recently been developed by Edwards (1) and McClelland *et al.* (2). Edwards' measure of this variable is one of fifteen that have their origin in, and bear the same names as, a list of manifest needs presented by Murray (3). They form the Edwards Personal Preference Schedule which measures a "number of relatively independent normal personality variables" (1, p. 1). McClelland's method of measuring n Achievement is also rooted in Murray, since he uses imaginative stories evoked by two TAT cards and two other pictures designed by him and his associates. He has developed a quantitative method of scoring the imaginative material in terms of n Achievement categories (2, pp. 107–138).

Since both Edwards and McClelland independently measure the same variable, a high correlation is expected between the scores obtained on their two measures. The purpose of this study is to test whether such a relation exists or not.

Method

The subjects (*S*s) in this experiment were 84 Arabic-speaking students in two psychology classes at the American University of Beirut. Approximately half of the *S*s were citizens of Lebanon, and the other half came from Jordan, Syria, Iraq, and Bahrain. All the *S*s had a good command of English, which is the language of instruction in the University. Sixty of the *S*s were males.

The full PPS was administered to the *S*s as a class demonstration in the use of the questionnaire method. The *S*s were told that *E* was interested in collecting norms on university students for help in his counseling work and were also told that these norms would be available to the class members. The *S*s were told to use a private identification mark on their answer sheets and were asked to keep a record of it. A week later the four pictures used by McClelland were projected on a screen, and each *S* was given four sheets of paper, on each of which were printed the four standard questions used by McClelland about the pictures which the *S*s were asked to answer. The instructions and method used by McClelland (2, p. 98) were followed throughout. The *S*s were asked to use the same identification mark they had used on the PPS the week before. This was also presented to the class as a demonstration of projective tests which were to be discussed during the following class session. Both these tests were given under similar conditions, and the situation approximates McClelland's "neutral" condition of testing n Achievement.

The two tests were scored according to the methods prescribed by their authors. Fifteen scores were available for each subject on the PPS. The imaginative stories of the 84 *S*s were independently scored by two research assistants and the interscorer reliability correlation was .93, which is considered to be satisfactory. Only the 69 *S*s, 50 males and 19 females, who completed both tests were included in the final analysis. The age range for the 68 *S*s who had indicated their age (on their last birthday) was 18 to 29, with a mean of 21.1 and a standard deviation of 2.2.

---

[1] The author wishes to acknowledge his indebtedness to Aimee Sabbagh and Sumaya Kurani, of the Department of Psychology, for scoring the imaginative stories on the McClelland Test.

Results and Discussion

Table 1 shows the means and standard deviations on Achievement for our *S*s and those of Edwards. The difference between the two means is not significant, *t* being .52, which points to the probable validity of this questionnaire for a non-American culture. To verify this, the Achievement items on the PPS were presented to three Arab professors in the fields of psychology, Arabic history, and Arabic literature. They were asked to judge whether any of the items were in any way alien to the Arabic culture and whether they could be taken as indices of achievement for the Arab group. There was full agreement among the judges about the validity of these items for an Arabic-speaking culture.

The mean score of the *S*s on the imaginative stories was 2.22 with a standard deviation of 4.26. No comparable data are available for American groups. Even though such information is not available, a glance at the distribution of the different categories of achievement scores in Table 2 points to the possibility of cultural differences on certain categories. These differences will be discussed in a separate paper as they are not directly related to the present study. Evidence regarding the cross-cultural validity of the scoring method used is presented by McClelland (2, pp. 168–172).

The correlation between the scores of the 69 *S*s on the PPS achievement variable and the McClelland achievement scores is .16. Since an *r* of .16 is not significant, it indicates an absence of a relationship between two methods that claim to measure the same variable.

Like all motives, the achievement motive is rather complex. McClelland (2, p. 217) describes his score as that of a "composite index of a variety of types of achievement motivation" which he demonstrates as having a meaningful relationship to varied types of behavior. Among the theoretical distinctions which he makes between different types of achievement is one type that is characterized chiefly by a "hope of success" and another by a "fear of failure" (2, p. 215).

Table 1

Mean Achievement Scores of Arab and American[a] Subjects on the PPS

Group	*N*	Mean	*SD*
Arabs	84	14.65	4.26
Americans	1509	14.38	4.36

[a] U. S. scores taken from Edwards (1, p. 6).

Table 2

Frequencies of Various Achievement Categories[a] under Neutral Experimental Conditions for Both Arab and American Subjects[b]

Achievement category	Americans	Arabs
Achievement Imagery (AI)	75	70
Doubtful Imagery (TI)	49	102
Unrelated Imagery (UI)	49	112
Need (N)	35	25
Instrumental Activity (I)	31	33
Positive Anticipatory Goal State (Ga +)	25	14
Negative Anticipatory Goal State (Ga −)	20	12
Personal Obstacle (Bp)	8	8
Environment Obstacle (Bw)	17	10
Positive Affect State (G +)	17	16
Negative Affect State (G −)	23	13
Nurturant (Nup)	5	5
Achievement Theme (Ach. Th.)	62	58

[a] The total number of stories for the Arab group is 276 (69 *S*s, 4 stories each); and for the American group is 156 (39 *S*s, 4 stories each).

[b] Data on the American *S* is taken from McClelland (2, p. 141).

This theoretical distinction may possibly account for the absence of a relationship between the two measures. An inspection of the PPS items which measure achievement motivation shows that they refer to competition with some standard of excellence similar to the "hope for success" type of achievement. "To do one's best, to be successful, to accomplish . . . to do . . . , to solve . . . , to do things better than others . . . etc. . . ." are given by Edwards (1, p. 5) in describing his Achievement variable.

Besides the above, responses to the Edwards items are made to clearly structured stimuli, whereas the imaginative stories are responses to a variety of cues both internal and external, in the environment and the pictures, as well as to experimentally induced cues. This difference points to the probability that the two measures tap different levels of achievement, a manifest level, and a latent level (2, pp. 196–217), and may account for the absence of a relationship between the two measures.

In the light of the findings, and in the interest of avoiding confusion, it is suggested that the two measures of achievement motivation be either renamed or qualified.

Summary

The PPS and the McClelland pictures for measuring achievement were given a week apart to a group of 84 Arabic-speaking *S*s at the American University of Beirut. No relationship was found between the achievement scores of the 69 *S*s who completed both tests. The results were discussed in terms of the complexity of the achievement motive and, also, of the probability that the two methods measure the achievement motive at different levels.

Received August 2, 1957.

References

1. Edwards, A. L. *Manual: Edwards Personal Preference Schedule.* New York: Psychological Corp., 1954.
2. McClelland, D., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.
3. Murray, H. A., *et al. Exploration in personality.* New York: Oxford Univer. Press, 1938.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

Second-Order Factors in Personality Measurement

Samuel Karson
University of New Hampshire

and Kenneth B. Pool[1, 2]
University of California, Berkeley

It is surprising that to date so few published studies on the Sixteen Personality Factor Questionnaire have appeared which have not emanated from the Laboratory of Personality Assessment at the University of Illinois. The reasons for this lack of popularity are not readily apparent, since Cattell and his coworkers have published widely and frequently (3, 4, 5, 7, 8). Karson and Pool (11) found much similarity between the 16 P.F. and the MMPI which they interpreted as offering support for the construct validity of the 16 P.F. test. Their major criticism of the test was that the coefficients of equivalence of the factors on Forms A and B were not nearly high enough to justify their being used as equivalent forms, and that even after being corrected for full length with the Spearman-Brown prophecy formula, the resultant coefficients of half of the 16 factors were still not high enough to warrant their use for accurate individual personality description.

In a recent paper, Cattell (5) reported four second-order personality factors which he identified as Anxiety-vs.-Dynamic Integration, Extraversion-vs.-Introversion, Cyclothyme-vs.-Schizothyme Constitution, and Unbroken Success-vs.-Frustration. These findings were based on a sample of 408 normal subjects composed of 181 undergraduate students of both sexes and 227 U.S. Air Force young men in training. The aim of the present study was to attempt to replicate the second-order factors described by Cattell in order to determine the degree of stability of these factors over different samples. From a theoretical viewpoint, if different samples resulted in yielding unrelated factor structures, then the argument (4, p. 3) that factors actually correspond "to psychological functional unities rather than to mere mathematical conveniences" would be highly questionable.

Sample

The sample consisted of 71 USAF officers who were referred to the School of Aviation Medicine for medical and psychological evaluation. The average age was 33, with a range of 20 to 39 years, and the mean number of years of education was 14, with a range of 12 to 16 years. The mean Full Scale IQ on the Wechsler Adult Intelligence Scale was 121.98, with a standard deviation of 6.50. The patients were referred for evaluation between August, 1955 and July, 1956. The sample employed in this research is believed to be typical of the flying personnel who are referred for psychiatric consultation in the U.S. Air Force.

None of the patients was psychotic, but almost all were manifesting some disturbance in functioning efficiency which was impairing, at least temporarily, their ability to perform their duties as USAF flying personnel. Upon clinical examination, which included psychiatric and psychological evaluation, there was general agreement that anxiety and its disrupting effects usually played a key role in the etiology of the patients' symptomatology,

---

[1] The data for this study were collected and the statistics computed while both authors were at the School of Aviation Medicine, USAF.

[2] A summarized version of this paper was presented at the 1957 meetings of the APA in New York City by the senior author.

Table 1

Intercorrelations* of Factor Scores on the Sixteen P.F. Test Forms A and B Combined ($N = 71$)

Score	*A*	*B*	*C*	*E*	*F*	*G*	*H*	*I*	*L*	*M*	*N*	*O*	Q_1	Q_2	Q_3	Q_4
A		.02	.12	.13	.42	−.10	.45	−.12	−.07	−.08	.12	−.07	.01	−.27	−.08	.11
B			−.12	.08	.00	−.08	−.11	.02	.21	.18	−.13	.21	−.15	.02	−.15	.20
C				.22	.30	−.17	.41	−.38	−.59	−.57	.03	−.71	−.40	−.13	.43	−.64
E					.41	−.03	.49	−.43	.00	−.07	.13	−.32	.12	−.01	−.13	−.08
F						−.20	.66	−.23	−.14	−.14	.12	−.28	−.02	−.16	−.19	−.13
G							.08	.16	.23	.03	−.07	.00	.07	.26	.17	.04
H								−.26	−.23	−.27	.18	−.52	−.17	−.36	.05	−.28
I									.20	.45	−.23	.40	.04	.16	−.09	.24
L										.52	.00	.46	.22	.22	−.18	.51
M											−.13	.56	.31	.29	−.36	.53
N												−.18	.11	−.01	−.03	−.12
O													.21	.23	−.41	.76
Q_1														.24	−.16	.17
Q_2															.01	.09
Q_3																−.42
Q_4																

* For significance at the .05 level of confidence a Pearson *r* must equal .24.

even though an actual incapacitating, diagnosable neurosis was seldom present. It should be remembered that the sample was composed of mentally and physically superior experienced military flying officers, many of whom had successful combat records in either World War II or Korea.

Method and Results

The 16 P.F. test (Forms A and B) was administered as a part of the routine diagnostic psychological test battery. The two forms of the test were alternated so that about one-half of the sample took Form A first, while the other half had Form B initially. Both forms of the test were usually administered on the same day, one in the morning and the other in the afternoon. The tests were scored according to the directions given in the handbook. The raw scores were converted into standard scores by using the sample of male college students and high school seniors whose average age was 21 as the reference group. Pearson coefficients of correlation among the standard scores for the 16 factors were computed to achieve the matrix of intercorrelations presented in Table 1.[3]

The factor analysis was then done, using Thurstone's centroid method (12). As guides for when to stop factoring, criteria suggested by Burt and Banks (1) and Guilford and Lacey (10) were used. After six factors had been extracted, an estimated 109% of the common factor variance had been accounted for, and the sixth factor residuals ranged in size from − .068 to .093. Rotation to simple structure was then attempted without regard to positive manifold. As a guide in rotation, the goal was to maximize the number of variables with zero projections on the reference vectors. The factors were rotated blindly two at a time by means of graphs and six independent, second-order factors were extracted.[4] The new factor loadings and their communali-

---

[3] Although this procedure for obtaining the matrix of intercorrelations is admittedly not as precise as when one begins with the cosine matrix among the simple structure primary factors and then inverts this matrix to obtain the correlations among the factors, Cattell has previously employed this method (4, p. 18). On this same issue, Cattell (5, p. 412) states: "However, it must be granted that where factors are already very purely measured by existing tests, with no systematic bias and with effective suppression of peer factors in each factor score, the factoring of actual factor scores, i.e., of one variable for each factor, may yield a reasonable approximation to second-order factors. This seems to have happened with the 16 P.F. test in the large Air Force study."

[4] The centroid factor matrix, the matrix of sixth factor residuals and the direction cosines have been deposited with the American Documentation Institute. Order Document No. 5586, remitting $1.25 for microfilm or $1.25 for photocopies.

Table 2

Rotated Second-Order Factor Loadings

First order factor	I	II	III	IV	V	VI	h^2
A	.03	.51	.09	.34	.10	−.02	.3951
B	.24	.03	.38	−.10	.03	−.07	.2187
C	−.84	.24	.27	−.17	.05	−.08	.8739
E	−.04	.54	−.02	−.33	.04	−.37	.5410
F	−.11	.75	.04	−.02	.20	−.03	.6175
G	.03	−.06	−.08	.01	−.61	.08	.3895
H	−.33	.84	−.04	.19	−.16	.00	.8778
I	.32	−.23	−.01	.03	−.09	.60	.5244
L	.65	−.02	.00	−.08	−.41	−.08	.6038
M	.72	.02	.00	−.23	−.04	.32	.6757
N	−.07	.11	−.30	.07	.01	−.27	.1849
O	.81	−.34	.18	.12	.09	.09	.8347
Q_1	.34	.02	−.47	−.17	−.02	−.05	.3687
Q_2	.21	−.23	−.05	−.47	−.26	−.01	.3881
Q_3	−.54	−.19	.03	.01	−.34	.04	.4459
Q_4	.80	−.06	.21	.19	−.04	−.04	.7270

ties after rotation of axes are presented in Table 2.

Description of the Factors

The chief indications of the nature of Factors 1 through 6 are presented below with a contingent title given to each factor.

Factor 1. Anxiety-vs.-Dynamic Integration

Symbol	*16 P.F. Handbook Title*	*Loading*
C	Ego Strength	−.84
O	Anxious Insecurity	.81
Q_4	Nervous Tension	.80
M	Hysteric Unconcern	.72
L	Paranoid Schizothymia	.65
Q_3	Will Control & Character Stability	−.54
Q_1	Radicalism	.34
H	Adventurous Autonomic Resilience	−.33
I	Emotional Sensitivity	.32

This factor shows a marked resemblance to the second-order factor reported by Cattell (5) on which Q_3, *O*, Q_4, *L*, *C*, and *H* loaded heavily. He identified this factor as Anxiety-vs.-Dynamic Integration and we have adopted his title. Further support for this identification is forthcoming from an earlier study by Cattell (4) in which a factor described by him as Anxiety-vs.-Self-Assuredness was reported. The 16 P.F. test factors, Q_4, *O*, Q_3, *C*, *L*, *M*, and *H* all loaded on this factor.

Additional support is provided by Karson and Pool (11) who reported that the MMPI psychasthenia (*Pt*) scale correlated significantly with factors *C*, *O*, Q_4, *M*, *L*, Q_3, *H*, and *I*. Moreover, the factors with significant loadings on Factor 1 correlated more highly with the *Pt* scale than any of the other 13 MMPI scales, with the possible exception of *Sc* which correlated .89 with *Pt* for the present sample (11). High correlations between *Pt* and the Taylor Anxiety scale have been reported by Eriksen and Davids (9) and by Brackbill and Little (2).

Factor 2. Extraversion-vs.-Introversion

Symbol	*16 P.F. Handbook Title*	*Loading*
H	Adventurous Autonomic Resilience	.84
F	Surgency	.75
E	Dominance	.54
A	Cyclothymia	.51
O	Anxious Insecurity	−.34

Factor 2 also shows much similarity to another of Cattell's second-order factors on which *M*, *F*, *A*, Q_1, Q_2, and *H* had significant loadings. He identified this factor as Extraversion-vs.-Introversion, and once again we have adopted his title (5). Cattell reports that in still another study (5, p. 416) the *E* factor (which we found) also enters in and, moreover, that an alternative rotation which was considered for his Factor 2 on 4 would also have done so again. The reasons why Q_1 did not load on Factor 2 in the present study, while *O* did, are unknown, although it is believed that sampling differences may be responsible for these variations.

Karson and Pool (11) provide additional support for the identification of Factor 2 inasmuch as they reported significant correlations between certain 16 P.F. scores and particular MMPI scale scores which appear relevant to this identification.

The present findings do not, however, support Cattell's assertion (5, p. 416) that the "relatively low loadings of the more hereditarily determined *H* and *A* factors suggest that Factor 2 is an environmentally determined pattern." Factor *H* had the highest loading of all in the present study. Nevertheless, his conclusion (5, p. 417) that "There now seems ground for recognizing extraversion-introversion as something more than a correlation cluster or surface trait, and perhaps for specializing the term for this measurable second-order factor" is in agreement with the present findings.

Factor 3. Unbroken Success-vs.-Frustration

Symbol	*16 P.F. Handbook Title*	*Loading*
Q_1	Conservatism	$-.47$
B	General Intelligence	.38
N	Lack of Sophistication (Naivete)	$-.30$
C	Ego Strength	.27

Factor 3 shows some resemblance to Cattell's factor of Unbroken Success-vs.-Frustration inasmuch as Q_1 and N are present in both patterns as well as high intelligence from an objective behavior study (5, p. 417).

Karson and Pool (11) demonstrated, however, that for the present USAF officer sample, Factor B was not a valid measure of intelligence. They also reported significant correlations between B and *Mf* and L, respectively, and between Q_1 and F, *Hy* and *Mf* from the MMPI. This pattern of relationships makes Factor 3 difficult to identify. The present authors have chosen to adopt Cattell's title of Unbroken Success-vs.-Frustration, although a rival hypothesis of Spontaneity-vs.-Inhibition was also entertained.

Factor 4. Independence-vs.-Dependence

Symbol	*16 P.F. Handbook Title*	*Loading*
Q_2	Lack of Resolution	$-.47$
A	Cyclothymia	.34
E	Lack of Dominance (Submission)	$-.33$

Factor 4 has some similarity to a second-order factor (4, p. 40) of Schizothyme Tenacity-vs.-Cyclothymia on which Factors E, F, H, A, and Q_2 loaded. This identification is difficult to reconcile, however, with a more recent second-order factor again identified by Cattell (5, p. 417) as Cyclothyme-vs.-Schizothyme Constitution on which Factors A, I, N, Q_3, and C loaded significantly.

Karson and Pool (11) reported significant correlations between Q_2 and *Si* and *Mf*, between A and *Si*, and between E and *Mf*, *Si*, and *Sc*, respectively. In view of this pattern of interrelationships, Factor 4 has been tentatively identified as Independence-vs.-Dependence. The handbook (8) for the 16 P.F. lends support for this hypothesis by its description of the $Q_2 -$ person (8, p. 11), the $A -$ and $A +$ person (8, p. 6), and Factor E as well (8, p. 7).

Factor 5. Sociopathic Deviancy

Symbol	*16 P.F. Handbook Title*	*Loading*
G	Lack of Super-Ego Strength	$-.61$
L	Lack of Paranoid Schizothymia	$-.41$
Q_3	Lack of Will Control	$-.34$
Q_2	Lack of Resolution	$-.26$

The identification of this factor is especially difficult, principally because the two factors with the highest loadings, namely G and L, are among the ones about which least is known. Karson and Pool (11) reported that none of the MMPI scale scores correlated significantly with Factor G which, strangely enough, had its highest correlations with *Pd* and K. These findings as well as the authors' clinical experience with this test over a two-year period raise some doubt as to the identification of G as a factor of "effective" superego pattern (6). Factor L is also puzzling; it did not correlate with *Pa;* however, it correlated more highly with *Pt* and *Sc* than did *Pa*. Moreover, the high loading of L on the second-order factor of Anxiety-vs.-Dynamic Integration, which has been found consistently (4, 5, 7), does not appear to have been satisfactorily explained on a theoretical basis although such an explanation has been attempted (7, p. 230).

Two hypotheses are suggested for the identification of Factor 5; first, that it is a factor of sociopathic deviancy and, second, that it represents a lack of obsessive-compulsive personality characteristics in our sample. The latter interpretation would necessitate a reexamination of Factor L with emphasis given to its anxiety and rigidity aspects rather than to paranoid suspicion. At any rate, it is encouraging to see some relationships emerge between the highly esoteric world of factor analysis and the more familiar conceptualizations of psychiatric nosology.

Factor 6. Latent Homosexuality (Passive)

Symbol	*16 P.F. Handbook Title*	*Loading*
I	Emotional Sensitivity	.60
E	Lack of Dominance (Submission)	$-.37$
M	Hysteric Unconcern	.32
N	Lack of Sophistication (Naivete)	$-.27$

The particular constellation of loadings on this factor suggests the hypothesis of latent homosexuality of a passive nature. Some support for this identification comes from the significant correlation reported between *I* and *Mf* (11) and from the description of the *I* + person presented in the 16 P.F. handbook (8, p. 9). This factor may serve to throw light on the psychodynamics of military flyers, particularly with regard to the frequently made clinical observation of "masculine strivings" for this group.

Some mention should be made of the fact that Cattell (5) used an oblique solution in obtaining the four second-order factors which he described, while the present study reports six independent factors. These differences notwithstanding, the present findings appear to be encouraging for those who believe in the stability and reality of factorial solutions. Why two of Cattell's factors correlated .43, while the same two factors were not significantly correlated in the present study, is not readily apparent although this discrepancy may be a function of methodological differences, sampling differences, or both.

Summary

In order to evaluate the stability of factorial solutions over different samples, an attempt was made to replicate the second-order personality factors reported by Cattell. Forms A and B of the 16 P.F. were administered to a sample of 71 USAF officers on flying status, who had been referred to the School of Aviation Medicine, USAF, for medical and psychological evaluation. The test scores were punched on IBM cards, and a correlation matrix was derived using Pearson product-moment correlation coefficients. The factor analysis was done using Thurstone's centroid method. Rotation to simple structure by means of graphs yielded six orthogonal, second-order personality factors. These were described as follows: 1. Anxiety-vs.-Dynamic Integration; 2. Extraversion-vs.-Introversion; 3. Unbroken Success-vs.-Frustration; 4. Independence-vs.-Dependence; 5. Sociopathic Deviancy; 6. Latent Homosexuality (Passive). The first three factors named above appear to be replications of three of the four second-order factors reported by Cattell and his titles for them were used. The remaining differences in the nature and number of second-order factors found between the two studies can probably be explained at least partially by methodological differences, by the varying degree of simple structure actually attained, and by differences in age, status, and sex as well as in the personality characteristics of the two samples.

Received August 5, 1957.

References

1. Burt, C., & Banks, Charlotte. A factor analysis of body measurement for British adult males. *Ann. Eugen., Camb.*, 1947, **13**, 238–256.
2. Brackbill, G., & Little, K. B. MMPI correlates of the Taylor Scale of Manifest Anxiety. *J. consult. Psychol.*, 1954, **18**, 433–436.
3. Cattell, R. B. *Personality: A systematic theoretical and factual study*. New York: McGraw-Hill, 1950.
4. Cattell, R. B. Psychiatric screening of flying personnel. *Personality structure in objective tests —A study of 1,000 Air Force students in basic pilot training. USAF Sch. of Aviat. Med.*, Proj. No. 21-0202-0007, Feb., 1955, 1–50 (Rep. No. 9).
5. Cattell, R. B. Second-order personality factors in the questionnaire realm. *J. consult. Psychol.*, 1956, **20**, 411–421.
6. Cattell, R. B. *Handbook for the I.P.A.T. Anxiety Scale*. Champaign, Ill.: Institute for Personality and Ability Testing, 1957.
7. Cattell, R. B. The conceptual and test distinction of neuroticism and anxiety. *J. clin. Psychol.*, 1957, **13**, 221–233.
8. Cattell, R. B., Saunders, D. R., & Stice, G. F. *The Sixteen Personality Factor Questionnaire*. Champaign, Ill.: Institute for Personality and Ability Testing, 1949.
9. Eriksen, C. W., & Davids, A. The meaning and clinical validity of the Taylor Anxiety Scale and the hysteria-psychasthenia scales from the MMPI. *J. abnorm. soc. Psychol.*, 1955, **50**, 135–137.
10. Guilford, J. P., & Lacey, J. I. *Printed classification tests*. Army Air Forces Aviat. Psychol. Progr. Res. Rep. No. 5. Washington, D. C.: U. S. Gov't Printing Office, 1947.
11. Karson, S., & Pool, K. B. The construct validity of the Sixteen Personality Factors Test. *J. clin. Psychol.*, 1957, **13**, 245–252.
12. Thurstone, L. L. *Multiple factor analysis*. Chicago: Univer. Chicago Press, 1947.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

The MMPI Performance of Veterans with Organic and Psychiatric Disabilities[1]

Joseph J. Motto
Detroit Regional Office, Veterans Administration

The purpose of this paper was to investigate the Minnesota Multiphasic Personality Inventory (MMPI) performance of two groups of disabled veterans to determine whether there were areas of significant difference in terms of personality functioning. Particular attention was devoted to the *K* correction scale to determine if significant differences existed in the "test-taking attitude" of veterans with psychiatric as compared to those of organic disabilities.

The setting of the study was a VA counseling service for seriously disabled veterans. The *S*s consisted of a group of 50 veterans having VA ratings for tuberculosis and gunshot wounds and a like number of veterans with VA ratings for schizophrenia. The *S*s were selected at random from the population of male counseling cases seen within the past three years.

The MMPI scores of the psychiatric (P) group were compared with the organic (O) group, and both of these groups were compared with the MMPI standardization group on several of the scales, and with a group of nondisabled "normal" veteran controls.

Examination and statistical analyses of the data revealed the following:

1. MMPI scales, *Pt* and *Sc*, significantly differentiated those of organic disabilities from those of schizophrenic disabilities.

2. In terms of personality characteristics measured by the MMPI, the organic group was more similar to the psychiatric group than they were to a group of nondisabled, "normal" veteran controls.

3. The presence of an organic disability resulted in *Hs* scores which were not significantly higher than those produced by the psychiatric group.

4. The *K* scale functioned as predicted by test authors and conformed to their rationale of the suppressor variable when applied to those with organic disabilities, the sole exception being the *Mf* scale where it appears a *K* correction weight is needed for this group. A similar trend to an inverse relationship between *K* and *Mf* was seen in the P group.

5. There were indications of the need for an additional *K* correction weight for veterans with psychiatric disabilities on *Sc*, *Ma*, and *Pa* scales.

6. Compared to the control group, both disability groups produced higher *F* scores. The organics differed from the normals by the presence of an elevated neurotic triad, whereas the psychiatric group differed from the normals by elevations of both the neurotic and the psychotic triads.

7. *K* was found to have a high positive relationship with *L* and a significant inverse relationship with *F*.

Brief Report.
Received March 10, 1958.

---

[1] An extended report of this study may be obtained without charge from Joseph J. Motto, 8635 Elmira, Detroit 4, Michigan, or for a fee from the American Documentation Institute. Order Document No. 5579, remitting $1.75 for microfilm or $2.50 for photocopies.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

Two New Experimental Scales of the MMPI

Kenneth B. Little[1]
Stanford University

and Jerome Fisher
VA Hospital, San Francisco, California

Several investigators (1, 2, 3) have made use of one or both of two experimental scales derived by us in the course of a cluster analysis of the Hysteria scale of the MMPI. The purpose of this note is to describe briefly the development of these scales and to indicate some of their correlates in clinical and experimental work.

As part of the study of differential personality patterns in diagnostic center and general medical patients in a Veterans Administration hospital, it was observed that the Hysteria scale score was the highest point on the MMPI profiles of both groups. In both, the mean T score exceeded 70 despite the fact that the general medical patients were hospitalized for nonpsychiatric disorders. The observation of an elevated *Hy* score in general medical patients is, of course, a common one. On the assumption that these elevated scores had a different meaning in a nonpsychiatric population than in one with some history of psychologic disorder, we initiated a cluster analysis of the matrix of tetrachoric intercorrelations among the 60 Hysteria scale items. The patient population on which the correlations were derived consisted of 60 diagnostic center neuropsychiatric subjects and 90 random admissions to the general medical wards.

Nine relatively clear-cut clusters were obtained. Five of these clusters had moderate to high positive correlations among themselves and zero or slightly negative correlations with the remaining four. The latter group of clusters had, conversely, moderate to high positive correlations within itself. The findings suggested, of course, that two primary dimensions underlay the cluster structure. The group of five clusters (comprising 23 of the *Hy* items) was composed for the most part of statements which consist largely of physiologic symptoms to which the subjects "admitted." This group was labeled the Admission (*Ad*) scale.

The second scale was composed of the 19 items falling into the group of four clusters. These items consisted for the most part of items scored if the subject denied their validity for him. In content, they were statements about poor interpersonal relations, feelings of hostility, suspicion, and the like. The second cluster was therefore called the Denial (*Dn*) scale. As a final step, biserial correlations were computed between each of the remaining 18 Hysteria scale items and the *Ad* and *Dn* scale scores. Items were added to the scale with which they had the highest correlation. Two items correlated zero with both scales and were discarded.

The item composition of the two scales (as they are numbered in the group form booklet of the MMPI) is as follows: *Ad* (32 items)—Items 2, 3, 7, 8, 9, 55, 103, 107, 128, 137, 153, 160, 163, 174, 175, 188, 190, 192, 230, 243 scored if marked *False*, and items 10, 23, 32, 43, 44, 47, 76, 114, 179, 186, 189, and 238 scored if marked *True*. *Dn* (26 items)—Items 6, 12, 26, 30, 71, 89, 93, 109, 124, 129, 136, 141, 147, 162, 170, 172, 180, 201, 213, 234, 265, 267, 279, 289, 292 scored if marked *False*, and item 253 scored if marked *True*. All 58 items are scored in the same direction as they are scored on the *Hy* scale.

---

[1] Now with the Training and Standards Branch, National Institute of Mental Health, Bethesda, Maryland.

Correlations between the new scales and other MMPI scales for two different groups of subjects,[2] $N = 100$ in each group, are given in Table 1. The Kuder-Richardson reliability coefficients based on all 200 cases were $Ad = .83$ and $Dn = .75$. *Ad* has its highest correlation with the MMPI *Hs* scale, a not too surprising observation in view of the 18-item overlap between the two scales. In most populations, *Hs* and *Ad* could be substituted one for the other. The *Dn* scale has as its highest correlation that with the MMPI *K* scale, with which it shares nine items in common.

In the clinical application of these scales, it has been found that high scores on the *Dn* scale generally describe the uninsightful, the anti-intraceptive, the morally virtuous individual; high scores on *Dn* are also seen with "muted" or pseudo-normal profiles. Extreme scores on the *Ad* scale, on the other hand, depict the person in general psychologic distress who complains and is anxious about his somatic functions and who also reports disturbances in his object relationships. Such high scores on *Ad* are usually seen with overall elevated MMPI profiles. When both scales have moderate to high scores, equivalent to an elevation on the original *Hy* scale, there is the distinct clinical impression of the presence and operation of conversion neurotic dynamics in the behavior and history of the individual.

Table 1

Correlations of Admission and Denial Scales with MMPI Scales for Two Populations

MMPI scale	General medical patients		Patients referred to psychiatric service	
	Ad	*Dn*	*Ad*	*Dn*
L	−09	35	−10	41
F	52	−23	64	−54
K	−33	78	−57	88
Hs	89	−29	90	−42
D	59	−08	70	−41
Pd	43	−27	46	−37
Mf	17	01	17	−14
Pa	22	17	48	−23
Pt	57	−60	75	−71
Sc	53	−43	75	−65
Ma	19	−46	57	−59
Si	44	−49	55	−62
Ad	—	−30	—	−56
Dn	−30	—	−56	—

Received July 29, 1957.

References

1. Block, J., & Thomas, H. Is satisfaction with self a measure of adjustment? *J. abnorm. soc. Psychol.*, 1955, **51**, 254–259.
2. Gordon, J. Interpersonal predictions of repressors and sensitizers. *J. Pers.*, in press.
3. Truax, C. B. The repression response to implied failure as a function of the hysteria-psychasthenia index. *J. abnorm. soc. Psychol.*, 1957, **55**, 188–193.

---

[2] The subjects were tested as part of a project supported by a grant (M-1132) from the National Institute of Mental Health.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

The Parental Attitudes of Mothers of Schizophrenics[1]

Marvin Zuckerman
Institute of Psychiatric Research, Indiana University Medical Center, Indianapolis

Mary Oltean
Carter Memorial Hospital

and Irwin Monashkin
Purdue University

Psychoanalytic theorists such as Fromm-Reichman (3), Rosen (6), and Arieti (1) have hypothesized that schizophrenia has its origin in the early relationship between mother and child. They imply that the mother of a schizophrenic child has certain destructive types of attitudes toward children which she conveys to the child through her behavior and verbalizations. Unfortunately, we must rely on retrospective evidence bearing on this hypothesis. Tietze (9) interviewed a group of mothers of schizophrenics and characterized them as overanxious, obsessive and domineering. However, no control group was compared with these mothers, and one may question the objectivity of an open interview with a mother already identified as the mother of a schizophrenic.

The need for more objective techniques to assess parental attitudes encouraged other researchers to develop objective instruments. The typical parental attitude survey uses items which are stated in the form of generalities and clichés in order to make them sound socially acceptable. The parent is asked to state whether she mildly or strongly agrees or disagrees with these items. Freeman and Grayson (2) used the Shoben Parent-Child Attitude Survey in comparing 50 mothers of schizophrenics with 50 control mothers. They found the mothers of schizophrenics scored significantly higher on the Possessive and Ignoring scales (on the latter scale they accepted the .10 level of confidence) than the control mothers. Unfortunately, they did not control educational or socioeconomic differences between their groups. From the way in which the control group data was gathered, it is likely that differences existed in the educational levels of the groups. Furthermore, since they relied on volunteer mothers solicited by volunteer students, various selective factors could have determined the scores in the control group.

Mark (4), in a more carefully controlled study, found that a significant number of attitude items distinguished a group of 100 mothers of schizophrenics from 100 control mothers. However, McFarland (5) attempted to replicate this study with the same items and failed to find a significant number of items distinguishing the two groups.

The purpose of the present study was to retest the hypothesis that mothers of schizophrenics have more severe (controlling and rejecting) parental attitudes than mothers of normals.

Procedure

The Parental Attitude Research Instrument (PARI) was the main instrument used in this study. This test was developed by Schaefer and Bell (7), with items modeled after the types used by Mark (4) and Shoben (8). An attempt was made to develop a number of

---

[1] The authors are grateful to Donald Moore and Margaret Morgan for their permission in making the facilities of Carter Memorial Hospital available for this research. Harry Brittain assisted in the analysis of the data.

short, but internally consistent, scales measuring specific types of attitudes, to permit somewhat broader interpretations than can be made from *post hoc* item analyses. The test consists of 23 scales, each of which measures some attitude theoretically relevant to personality development in children. Normative data are furnished by Schaefer and Bell (7), and in a previous article by Zuckerman, Ribback, Monashkin, and Norton (11). A factor analysis of the scales using a heterogeneous sample, including the group reported in this paper, can be found in the latter article. The factors found in this study were used to derive summary scores in the current study by weighing scales roughly by their factor loadings. Factor A seems to be an Authoritarian-Controlling factor which loads heavily on 16 of the scales. Factor B seems to be a Hostility-Rejection factor which loads on three scales. Factor C seems to be a Democratic Attitude factor which loads on three scales.

Subjects

Forty-seven mothers of male and female patients diagnosed as schizophrenic in an acute treatment hospital were given the PARI. Ratings of pathology on 30 cases of mixed diagnoses were available from a previous study and provided some information on the reliability of diagnoses at the institution. A seven-point rating scale ranging from mildly neurotic to severely psychotic was used in rating the case-history summaries of the patients. Ratings were global, but factors considered included severity of symptoms, chronicity of the disorder, response to treatment, and premorbid adjustment. Two of the authors made the ratings independently, and their ratings correlated .80. The agreement between the raters and the diagnosis assigned the patient is summarized in Table 1. It can be seen that there is high agreement between psychiatric diagnosis and the ratings except in the middle of the scale where the borderline cases would be expected to fall. Although there is probably some contamination between the two sets of data, the experimenters felt justified in using the psychiatric diagnosis of schizophrenia as a criterion, recognizing that wide behavioral differences may exist within the disorder diagnosed as "schizophrenia."

Table 1

Relation Between Case History Rating and Psychiatric Diagnosis

Psychiatric diagnosis	Case history rating 1–3	3.5–4.5	5–7
Nonschizophrenic	8	2	1
Schizophrenic	0	2	17

Note.—Chi square, corrected (combining cells) = 15.56. $df = 1$, $p < .001$.

Table 2

Composition of Mothers of Schizophrenic and Mothers of Normal Groups

Level of education	Schizophrenics' mothers *N*	Schizophrenics' mothers Mean age	Normals' mothers *N*	Normals' mothers Mean age
Grammar school	11	52.4	11	52.0
High school	24	48.0	24	48.8
College	7	50.7	7	49.1
All cases	42	49.6	42	49.7

A group of 47 mothers of normals was derived from the older age range of the normal sample (11) and from some additional normals gathered at a later date. The control group consisted of mothers given the PARI at church social and nursery school meetings, with some mothers of students from a night extension college.

Since education of the mother seems to influence most of the scales and one of the factors (11), it was put into the analysis of variance design using three levels: grammar school, high school, and college. Age differences were then compared between diagnostic groups across levels. At the grammar school level, there was a significant age difference; therefore, the five oldest mothers of schizophrenics and the five youngest mothers of normals were discarded. The resultant samples, containing 42 mothers in each group, are described in Table 2. Sex of the patient was considered before grouping mothers of male and female schizophrenics together. Twenty-one mothers of male schizophrenics were compared with 21 mothers of female schizophrenics. A significant difference was found on only one scale. Therefore, the groups were combined. Other variables considered were the child's diagnostic subclassification, paranoid vs. other schizophrenics, and the time interval between the admission of the patient and the administration of the test to the mother. Neither of these variables seemed to markedly influence the average parental attitude scores.

Results and Discussion

The results of the analyses of variance on Factors A, B, and C, and on 20 of the individual scales are reported in Table 3. Only one of the differences between groups was significant. The mothers of schizophrenics scored significantly lower on Scale 8, strictness, than the controls. However, 1 out of 23 comparisons may be considered a finding attributable to chance. The differences between education levels were significant for Factor A and 14 of the 20 individual scales. The interaction effects were significant for Factor A and eight of the individual scales. An analysis of the interaction effect in Factor A can be seen in Table 4. At the grammar school level, the mothers of schizophrenics tend to score higher than the controls, and at the high school level they scored significantly lower than the controls. No differences were found at the college level. The results on the individual scales tend to follow the same pattern. If we had included the 10 subjects discarded to make age equivalent at the grammar school level, the interaction effect would have been even more pronounced.

The interaction effects are difficult to interpret. They may indicate that the way the

Table 3

Analysis of Variance Results in Comparisons of Mothers of Schizophrenics and Mothers of Normals[a]

Score or scale	Child's diagnosis	Mother's education	Interaction
Summary Scores			
A—Authoritarian-Control	—	15.41**	5.97**
B—Hostility-Rejection	—	1.89	—
C—Democratic Attitudes	1.39	—	—
Scales			
2—Fostering Dependency	—	5.80**	1.11
3—Seclusion of the Mother	—	8.80**	1.80
4—Breaking the Will	3.18	13.07**	3.98*
5—Martyrdom	—	12.21**	1.20
6—Fear of Harming the Baby	—	2.76	7.90**
7—Marital Conflict	—	3.99*	1.33
8—Strictness	6.31*	2.80	2.85
9—Irritability	—	1.44	—
10—Excluding Outside Influences	3.23	10.51**	3.71*
11—Deification	3.02	11.07**	4.57*
12—Suppression of Aggression	—	4.24*	2.02
13—Rejection of Homemaking Role	1.68	—	—
15—Approval of Activity	1.35	3.82	2.07
16—Avoidance of Communication	1.71	10.30**	1.01
17—Inconsiderateness of Husband	—	—	—
18—Suppression of Sex	—	11.75**	4.62*
19—Ascendance of the Mother	—	11.13**	6.12**
20—Intrusiveness	—	9.78**	2.69
22—Acceleration of Development	—	4.10*	4.31*
23—Dependency of the Mother	—	3.53*	3.62*

[a] F ratios of less than one are indicated by dashes.
* Significant at or below the .05 level.
** Significant at or below the .01 level.

Table 4

Comparisons of Means of Mothers of Schizophrenics and Mothers of Normals by Levels on Summary Score A

Level of education	Schizophrenics' mothers	Normals' mothers	t	p
Grammar school	242.09	214.91	1.88	.05–.10
High school	191.58	219.50	2.86	.01
College	172.86	155.86	.94	.3–.4

mothers of schizophrenics respond to the test depends on their familiarity with the cultural expectations for parental attitudes, which, in turn, may depend on their education. The samples at each level are not adequate to justify a conclusion. There is, however, a greater relationship between education and parental attitudes in mothers of schizophrenics than in the controls. The correlation between Factor A and education in the original 47 mothers of schizophrenics was − .632. In the original 47 controls it was only − .287. The difference between these two correlations was significant beyond the .05 level.

The hypothesis of a difference between parental attitudes of mothers of schizophrenics and other mothers was not confirmed. It would be difficult to disprove this hypothesis, since the possibility always remains that other instruments or techniques may reveal a difference. Toms (10) compared 30 mothers of male schizophrenics and 30 control mothers, using a variety of instruments including an interview, vocabulary test, the MMPI, and a short TAT. Unfortunately, there was a difference in education between the two groups which was not taken into account in the analyses. However, the general conclusion was that the hypotheses of personality differences between mothers of schizophrenics and other mothers were not confirmed from any of the techniques. One interesting result, which bears on our speculations about our interaction effects, was that the mothers of schizophrenics scored significantly higher on the two MMPI scales measuring defensiveness (L and K). These findings suggest that further research on the "schizophrenogenic mother" hypothesis might do better using tests which control for social desirability and response suppression.

Summary

It was hypothesized that mothers of schizophrenics have more severe (controlling and rejecting) parental attitudes than mothers of normals. Forty-two mothers of schizophrenic patients and 42 control mothers were compared on the Parental Attitude Research Instrument. The hypothesis was not supported, although significant interactions were found between the two groups and levels of education.

Received July 1, 1957.

References

1. Arieti, S. *Interpretation of schizophrenia.* New York: Robert Brunner, 1955.
2. Freeman, R. V., & Grayson, H. M. Maternal attitudes in schizophrenia. *J. abnorm. soc. Psychol.,* 1955, **50**, 45–52.
3. Fromm-Reichmann, Frieda. Notes on development of treatment of schizophrenics by psychoanalytic therapy. *Psychiatry,* 1948, **11**, 263.
4. Mark, J. C. The attitudes of the mothers of male schizophrenics toward child behavior. *J. abnorm. soc. Psychol.,* 1953, **48**, 185–189.
5. McFarland, R. L. Personal communication, 1957.
6. Rosen, J. N. The treatment of schizophrenic psychoses by direct analytic therapy. *Psychiat. Quart.,* 1947, **21**, 3–37.
7. Schaefer, E. S., & Bell, R. Q. Development of the Parental Attitude Research Instrument. *Child Develpm.,* in press.
8. Shoben, E. J., Jr. The assessment of parental attitudes in relation to child adjustment. *Genet. Psychol. Monogr.,* 1949, **39**, 101–148.
9. Tietze, T. A study of mothers of schizophrenic patients. *Psychiatry,* 1949, **12**, 55–65.
10. Toms, E. C. Personality characteristics of mothers of schizophrenic veterans. Unpublished doctoral dissertation, Univer. of Minnesota, 1955.
11. Zuckerman, M., Ribback, Beatrice B., Monashkin, I., & Norton, J. A. Normative data and factor analysis on the Parental Attitude Research Instrument. *J. consult. Psychol.,* 1958, **22**, 165–171.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

Representative Elements in Role Construct Technique

Spiro B. Mitsos
Evansville State Hospital

According to the psychology of personal constructs as set forth by Kelly (2), each individual has a conceptual system (an organized set of personal constructs) with which he gives structure to the world around him. Basic to the theory is the constant attempt by each individual to anticipate events in his surrounds. Hence, man is thought of in somewhat the same light as the scientist in his continual process of prediction and validation. Just as the scientist begins with a theory and makes his predictions on the basis of the concepts from that theory, man has a theory of life composed of bipolar personal constructs. His predictions are made along the lines of these personal constructs.

In applying this theoretical framework, Kelly and his associates have developed a technique designed to elicit the construct framework of any given individual. This technique is the Role Construct Repertory Test (RCRT).

With the RCRT, the subject (*S*) thinks about certain people he knows—people meaningful to him and who represent a variety of types. These people are referred to as "elements" and are determined by a role-title list. The *S* assigns people he knows to fulfill the role titles. Examples are as follows:

1. Write your own name in the first blank here.
8. Your closest present friend of the same sex as yourself.
17. The most successful person whom you know personally.

The *S* is then given these elements three at a time and asked to think about the three people and decide some important way in which two are alike and which sets them off from the third person. The contrasts are listed. Kelly assumes they are important constructs in the individual's personal construct system.

Certain other assumptions underlie the RCRT. One of these is the assumption of the "representativeness of the elements." "If the test is to indicate how the subject develops his role in the light of his understanding of other people, it is necessary that the other people appearing as elements in the test be sufficiently representative of all the people with whom the subject must relate his self-construed role" (2, p. 230). On the basis of this assumption, Kelly constructed a list of role titles; i.e., self, mother, father, ex-flame, pal, authorities, etc.

This assumption is the focus of concern in this paper. Kelly assumes representative elements (representative role titles) are necessary to elicit the most crucial personal constructs. The present research began with the simple question: "What would happen if the representative elements assumption were violated—and why?"

The problem was further limited to the question of consistency of construct labels with and without the fulfillment of the above assumption. Hunt (1) has previously investigated certain aspects of consistency with the RCRT. He examined *S*s on two occasions with a one-week time interval. Essentially different role-title lists were used each time. However, the two lists were equated as to sex, relatives, authority figures, etc. With his technique of determining equivalence of constructs, he found the average percentage of agreement in constructs used on the two protocols to be 69% with a standard deviation of 6%.

In Hunt's research, the assumption of representative elements was fulfilled in both forms of the test. In the present research, we are concerned with the question of whether or not similar results will occur if the assumption is violated in one form. The predictions were as follows: Construct labels elicited by the RCRT will be consistent through time when a representative role-title list is used. Consistency will also occur when the role-title list is nonrepresentative, in this case, homogeneous.

Procedure

The *S*s for this investigation included 20 psychiatric aides randomly divided into two equal groups. One of these groups is referred to as the "representative elements group (REG)"; the other, the "nonrepresentative elements group (NREG)." Constructs were elicited from the *S*s in the REG in the same manner as the Role Construct Repertory Test as described by Kelly. (In the present form there were 19 role titles and 22 sorts of 3 elements each.) All *S*s were re-examined with the same technique after a three-month interval. The *S*s in the NREG were treated in the same manner with a singular exception—the role-title list was not used. In its place, each *S* was merely asked to list the names of 19 personal friends.

According to the hypothesis, we would expect a significant number of constructs elicited on the first administration to be repeated on the second administration. Such would apply to both groups. Each construct for each *S* on the first administration was evaluated individually. If either pole of a given construct were repeated verbatim anywhere in the entire repertory elicited on the second administration, this was considered as repetition. Binomial expansion was utilized to establish 5% significance levels for number of constructs repeated and number of *S*s repeating a significant number of constructs.

Results

All protocols were evaluated by comparing both protocols for each *S*. If either or both poles of any given construct on the first administration were repeated verbatim on the second administration, a replication was scored. Thus the potential range of scores was 0–22. Applying the binomial expansion, 16 of 22 replications would be significant at the .05 level.

In the representative elements group, nine of the ten *S*s repeated significant numbers of constructs ($p < .02$). In the nonrepresentative elements group, two *S*s did so ($p > .20$). (One *S* in the latter group had to be dropped because he was not available for re-examination.) The results seem to indicate a direct relationship between representative role titles, on which to base constructs, and consistency of construct labels.

Table 1

Distribution of Replicated Constructs and Significant Subjects

	REG	NREG	χ^2	p
Constructs Repeated	170	122		
Constructs Not Repeated	50	76		
Significant Subjects	9	2	6.36	<.02
Nonsignificant Subjects	1	7		

Chi square was computed for the two groups comparing the number of *S*s repeating significant numbers of constructs with the number repeating insignificant numbers. These data are presented in Table 1.

Certain questions arose at this point of the research. Perhaps the most obvious was the possibility that these results are the result of the role-title list yielding the same people on each administration, while a different group of friends was used each time. Another question is that of whether or not people, when responding to such a test, are responding to the types of people described on the role-title list rather than or in addition to the particular people involved? Lastly, are different types of constructs elicited when the elements are heterogeneous as opposed to homogeneous?

In an attempt to clarify the results, the *S*s in the NREG were re-examined after another three-month interval. On this administration, Ss used the identical list of 19 friends they had produced on the second administration. These friends were listed in identical order, and the Ss were given identical sorts to elicit constructs. In comparing the second to the third administration, three *S*s repeated significant numbers of constructs ($p > .20$)—substantially the same results as the previous administration.

Discussion

It would seem apparent, on the basis of the above, that fulfillment of the assumption of the representativeness of the elements is a matter of importance, at least in terms of consistency. Construct labels were consistent through time when the assumption was fulfilled and were inconsistent through time when the assumption was violated.

It would appear, therefore, that conceptual consistency on the RCRT, at least insofar as it applies to the construct labels, must be in some way a function of the role-title list and the types of constructs elicited thereby.

The reasons can only be conjectured. The most likely source of explanation is Kelly's theory itself. The fundamental postulate of his system is stated as follows: "A person's processes are psychologically channelized by the way in which he anticipates events" (2, p. 46). One of the corrolaries of this postulate, the "Organization Corrolary," states: "Each person characteristically evolves, for his convenience in anticipating events, a construction system embracing ordinal relationships between constructs" (2, p. 56).

According to this corrolary, personal constructs within any individual's system are arranged in a hierarchical structure. "Within a construction system there may be many levels of ordinal relationships, with some constructs subsuming others and those, in turn, subsuming still others" (2, p. 57). Subsuming constructs are termed superordinal and subsumed constructs subordinal.

Relating these corrolaries to the present data, it would seem plausible that, when thinking of the people represented on the role title list, *S*s use personal constructs capable of describing not only specific individuals, but also describing similarities and differences between groups or types of people, i.e., family, authorities, peers, etc. Description in this case refers to the process of labeling dimensions on which one could anticipate the behavior of these individuals and groups.

In the present research, the "nonrepresentative elements group" described only a group of friends. These *S*s were not concerned with describing groups; their sample of people was fairly homogeneous. They were concerned, as it were, with only one possible subgroup of the heterogeneous group used by the "representative elements group." The construct of "friend versus nonfriend" might well be one of the group of dimensions used with the heterogeneous group of people.

In other words, it seems plausible that consistency occurs when the task requires the use of generally superordinate constructs. More specifically, it might be hypothesized that the number of superordinate constructs produced in a concept formation situation is a function of the heterogeneity of the objects to be construed. It may well be that each individual has an infinite number of subordinate constructs and hence a great deal of leeway in construing within the context of a superordinate construct; in this case, the context of friend versus nonfriend. Such an explanation would seem to be consistent with Kelly's fundamental postulate and the "Organization Corrolary."

Summary

The purpose of this study was essentially that of testing the differential effects of heterogeneity and homogeneity of role titles on the Role Construct Repertory Test. Two groups of subjects were administered the test. One group utilized a heterogeneous role-title list, the other a homogeneous group of elements. Significant differences were found in terms of the consistency of construct labels. In the group using the role-title list, a significant ($p < .02$) number of subjects demonstrated consistency of construct labels through time. In the group using homogeneous elements, only 20 per cent ($p > .20$) of the subjects demonstrated such consistency.

It would appear most likely that such results are a function of the heterogeneity of the role-title list. When presented with a heterogeneous group of stimulus objects on which to perform sorts and attach construct labels, individuals tend to produce more superordinate constructs, whereas more subordinate constructs are produced when the construed objects are homogeneous.

Received July 18, 1957.

References

1. Hunt, David E. Studies in role concept repertory: Conceptual consistency. Unpublished master's thesis, Ohio State Univer., 1951.
2. Kelly, George A. *The psychology of personal constructs.* New York: Norton, 1955.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

The Taylor A Scale and Clinical Measures of Anxiety[1,2]

Carl G. Lauterbach
Madigan Army Hospital

This study was designed to determine the relationships of the Taylor A Scale to clinical criteria of manifest anxiety, and to compare these relationships with those between Welsh's Anxiety Index (AI) and the same clinical criteria.

The *S*s were 44 male members of the U. S. Armed Forces referred to a military hospital for psychiatric consultation.

Three psychologists and a psychiatrist independently rated each *S* on a combined check list and rating scale of anxiety. The psychiatrist completed his ratings after his initial diagnostic interview. Then the psychologists simultaneously performed their ratings while one of them also administered the Wechsler-Bellevue Intelligence Scale to the *S*. Finally, the MMPI was administered, which provided both the A scale and AI variables.

The intercorrelations among the psychologists in their anxiety ratings ranged between .78 and .80. However, none of their ratings showed a significant relationship to the psychiatrist's ratings. Thus, a composite of the three psychologists' ratings was treated as a separate criterion of anxiety from the psychiatrist's ratings. An item analysis of the check list of anxiety signs resulted in four signs being selected, assigned weights, and summed for each *S* to form a third criterion of anxiety, referred to as the composite weighted check list of the psychologists. A psychiatric diagnosis formed the fourth criterion of anxiety.

The A scale correlated .44 ($p < .01$) with the composite rating of the psychologists and .38 ($p < .05$) with the composite weighted check list. The AI correlated .36 ($p < .05$) with the composite rating of the psychologists. Neither the A scale nor the AI correlated significantly with the psychiatric criteria of anxiety.

The two anxiety test measures and the four criteria of anxiety were correlated with the three validity scales of the MMPI and with the $F - K$ dissimulation index. The A scale related significantly to the F, K, and $F - K$ scales, while there was a lack of significant relationships, in general, between the validity scales and the criteria of anxiety. In the case of the K scale, it appears that K could be used as a variable to suppress the error in the A scale to improve the latter's relation to the criterion measures.

Brief Report.
Received March 15, 1958.

---

[1] This study is based upon a dissertation submitted for the Ph.D. degree, University of Iowa, 1952. The writer wishes to express his gratitude to Grant W. Dahlstrom, Committee Chairman, and the several staff members of the Neuropsychiatric Service, Fitzsimmons Army Hospital, who so freely participated in obtaining the data.

[2] An extended report of this study may be obtained without charge from Carl G. Lauterbach, 243 South 80th St., Tacoma, Wash., or for a fee from the American Documentation Institute. Order Document No. 5580, remitting $1.25 for microfilm or $1.25 for photocopies.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

Experience of Clinicians, and Speech Characteristics and Statements of Clients as Variables in Clinical Judgment[1]

Austin E. Grigg[2]
State University of Iowa

Whether he is reading a case history, conducting a psychological test, or interviewing a client, the clinical psychologist has as a major objective the formulation of certain kinds of judgments about his client. On the basis of these judgments, the psychologist develops his plans for the management of the case or for the kinds of recommendations he will make in a professional report. Very little is known, however, about the variables that are operating in making these judgments.

There appear to be some empirical findings to suggest that clinical experience, per se, does not insure more accurate clinical judgments about clients, as shown by Luft (9, 10), Cline (2), and Soskin (11). Some studies show that the method by which clinicians acquire information about the client may not have to be the traditional interview procedure in order for accurate judgments to be made, as, for example, the Borke and Fiske study (1) finding that clinicians were able to make as accurate predictions about neurotic patients after reading interview statements as they could make after face-to-face interactions with these patients, and Luft's report that readers of interviews did as well as listeners in predicting client responses to objective questionnaires. Nor is it clear today just how much information a clinician needs about a client in order to make valid judgments, or what kinds of data he needs. Stern, Stein, and Bloom (12), for example, found that increased observational and test data tended to be associated with more accurate judgments about individuals, but Gage (3) reported a lowering of judgmental accuracy with increased observational data and Kostlan (7) found that unless a social history was included, increased clinical testing data did not make for better predictions.

In an effort to learn more about the variables in an interview situation that may play a role in influencing the judgments of clinicians about their clients, three variables were manipulated in the present study: the voice of the interviewee, the content of the interview statements, and the training and experience of the clinician.

Procedure

The first counseling interviews of three male clients with personal problems who reported to a university counseling service were tape recorded. During this first visit, the clients indicated their appraisal of themselves by making Q sorts with Deck 10 from Hilden's *Random Sets of Personal Concepts* (6), by answering Gough's Adjective Check List (4), and a 25-item Self Report Questionnaire adapted from the Guilford-Martin Inventory of Factors GAMIN (5). After repeating these measures one week later, each of the clients was included in the study since the reliabilities of the Q sorts were .80 or higher, and the percentage of overlap of items on the other two measures exceeded 85 per cent.

From each of the clients' interviews, a ten-minute portion of the tape recording was selected on the basis of giving a general statement of the client's complaints without stress on specific etiological material, and an exact typescript was prepared. Then, an experienced actor re-enacted the interview selection with the original counselor in whatever manner the actor felt the script indicated, but without the actor hearing any recordings of the actual client. The same male actor was used to re-enact each script.

---

[1] This paper is based upon a doctoral dissertation completed at the State University of Iowa, August, 1957. The writer acknowledges the advice and suggestions of Leonard D. Goodstein, under whose direction this study was done. Gratitude is also expressed to Leonard Feldt for his consultation about statistical processing of the data.

[2] Now at the Department of Psychology and the Testing and Guidance Bureau, The University of Texas.

Thus, for each of the three clients, any one of three sets of data could be supplied to a given group of clinical judges: (*a*) the actual tape-recorded selection with the client, (*b*) the tape-recorded selection as re-enacted by the actor, or (*c*) an exact typescript of the interview selection. Also, since there were selections from interviews of three clients, it was possible to vary the content of the interviews.

Twenty-four male clinical psychologists with the Ph.D. degree and affiliation with the American Psychological Association were used as the Ph.D. judges. Twenty-four male trainees in clinical psychology constituted the Trainee judges and included either graduate students with at least one year of practicum experience, or clinical trainees in a hospital who had completed at least one year of graduate work in psychology. A group of Naive judges, 24 male undergraduate students from the introductory psychology course, were obtained from Coe College. The Ph.D. and Trainee judges were obtained from three Veterans Administration hospitals: Downey, Illinois, Iowa City, and Knoxville, Iowa; from the Elgin (Illinois) State Hospital, and from the Psychopathic Hospital and the Department of Psychology of the State University of Iowa.

Groups of judges were given various combinations of the voice and content conditions and after each presentation were asked to predict (actually, postdict) how the client responded to the three personality tests. The accuracy of any judge's predictions was measured by the number of agreements between these predictions and the responses of the clients, called "Client Key." Later, although they were instructed to predict the responses of the clients and not the impressions of other psychologists about the clients, the judges' predictions were scored by an "Expert Key" made from a consensus of three psychologists as to what the clients' responses "should be" after these psychologists had studied, independently, the complete case folder of each client.

Each of the three experience levels of judges was subdivided into three subgroups of eight judges, and the sequence of procedure was so arranged that any one subgroup heard a different voice and content combination than either of the other two subgroups. Each subgroup, but with different voice-content combinations, had the following three conditions: (*a*) a client speak his own interview content, (*b*) an actor re-enact a different one of the three interviews, and (*c*) an exact typescript of the third and different interview selection. The order of presentation was confounded with sequence subgroups, so that each of the above conditions occurred in each possible order of presentation as first, second, or last. By using a Lindquist Type IV analysis of variance design (8, pp. 285–288), the voice and the content conditions were "within" effects, and the experience levels were "between" effects.

Results

Table 1 presents the means and the standard deviations of the three variables—experience, voice, and content—as measured by the Client Key. A summary table of *F* ratios from

Table 1

Means, Standard Deviations of Prediction of Client Responses with Variables Manipulated in the Study

	Q Sort		SR		Gough ACL[a]	
	Mean	Sigma	Mean	Sigma	Mean	Sigma
Experience:						
Naive	.300	.063	14.46	1.51	37.29	18.68
Trainee	.368	.063	15.85	1.22	42.89	13.90
Ph.D.	.376	.066	16.43	1.17	42.39	20.87
Voice:						
Client	.344	.107	16.39	2.22	39.67	16.73
Actor	.333	.124	14.60	2.80	38.81	14.99
Script	.367	.108	15.75	2.18	44.10	16.49
Content:						
#1	.328	.112	15.25	2.44	51.72	14.92
#2	.366	.128	14.83	2.93	33.53	12.41
#3	.350	.099	16.65	1.68	37.32	15.23

[a] To reduce 300-item Gough ACL scores to smaller numbers, 190 was subtracted from total agreements between judge and client responses.

Table 2

Summary Table Showing *F* Values for the Three Measures

Key	*Q* Sort		Self Report		Gough ACL	
	Client	Expert	Client	Expert	Client	Expert
Between Effects						
Experience	9.55**	11.15**	13.76**	3.95*	2.47	0.01
Voice-Content (b)	1.02	3.39*	1.15	0.67	6.98**	3.68*
Voice-Cont.-Exp. (b)	0.53	2.61	1.08	0.88	2.09	0.31
Within Effects						
Voice	1.93	12.41**	15.02**	12.47**	5.32**	2.24
Content	2.35	54.62**	16.55**	22.50**	60.83**	16.93**
Voice-Content (w)	1.87	1.71	8.69**	10.12**	9.85**	0.72
Voice-Experience	0.56	1.74	0.79	0.71	0.38	0.98
Content-Experience	0.26	1.04	0.48	1.84	2.35	2.64
Voice-Cont.-Exp. (w)	1.68	0.36	1.55	0.04	0.68	0.51

* Significant beyond .05 level of confidence.
** Significant beyond .01 level of confidence.

the analysis of variance is presented in Table 2.

With the *Q* sorts, it can be observed that the two trained groups of judges were more accurate in their predictions than the naive group, and this difference was shown by further analysis to be statistically significant ($p < .01$), but the difference between the accuracy of predictions by Trainee judges and Ph.D. judges is not statistically significant. The differences between the three voice conditions are not significant with the *Q* sorts, although the actor's voice tended to lower the accuracy of the predictions about the client's responses. Also, with the *Q* sorts, when different interview contents were used, there were no statistically significant differences in the accuracy of the predictions.

With the Self Report Questionnaire (SR), both groups of trained judges were significantly more accurate in predicting the clients' responses than were the Naive judges ($p < .01$), but again the difference between the Trainee judges and the Ph.D. judges was not significant. With SR, the three voice conditions and also the three different contents of the interviews influenced the accuracy of the judges significantly, as shown in Table 2. Because there was a Voice-by-Content interaction, the effects of voice conditions were compared over each content separately and it was found that the client's voice produced the most accurate predictions for Content 1, but no significant voice effects occurred with the other content conditions with SR.

With the Gough Adjective Check List, although the two trained groups again appear to be superior to the Naive judges, statistical analysis summarized in Table 2 indicated that the differences are not statistically significant. Again, the actor's voice was associated with the poorest predictive accuracy. Content 1 was associated with accuracy scores strikingly superior to those obtained when Content 2 or Content 3 occurred ($p < .01$). Because with Gough ACL a Voice-by-Content interaction occurred, an analysis similar to that described above with SR data was performed, and it was observed that the client's voice was associated with significantly *lower* accuracy scores with Content 3, but this did not occur with the other contents.

With the Client Key, the three measuring instruments used—*Q* sorts, SR, and Gough ACL—reacted differently to the variables of the experiment (Table 1). With the *Q* sorts and the Gough ACL, the script gave the most favorable scores for the voice conditions. With each of the three personality measures, the actor's voice lowered the accuracy of the predictions. In regard to the content of the interviews, the three personality measures reacted differently: Content 1 gave strikingly superior scores over other contents with the Gough ACL, but was the poorest content condition with *Q* sorts. Content 2 was associated with the best scores on the *Q* sorts, but was the poorest content condition with the other two measures. With SR, Content 3 was associated with the highest predictive accuracy.

Table 3 presents the means and the standard deviations of the three variables as scored by the Expert Key. The three clinicians who composed the key, working independently, answered the three personality measures as they believed the client "should have responded" after they had studied all pertinent case history and interview reports as well as the actual recorded selections heard by the judges in the experiment. The average *Q* sort of the three judges and those SR and Gough ACL items on which two or more of the three "experts" agreed, constituted the Expert Key. As may be seen, the trends in the data are similar to those of the Client Key (compare Table 1 and Table 3), although now the accuracy scores are considerably higher. The major discrepancy between the results based on the two keys occurs in the *Q*-sort data. There, significant voice and content effects are revealed with the Expert Key, as shown in Table 2. In addition, the "between" component of the Voice-by-Content interaction was significant, and the latter effect also occurred with the Gough ACL data. Such a "between" effect may be associated with the disproportionate number of State University of Iowa judges in the three sequence subgroups. Subgroup I contained twelve such judges, whereas Subgroup II contained seven, and Subgroup III contained none. While this effect did not influence the predictions of how the clients responded, it apparently exerted an influence when the judges were scored by an Expert Key devised by three State University of Iowa psychologists.

Table 3

Means, Standard Deviations with the Variables Manipulated in the Study (Expert Key)

	Q Sort		SR		Gough ACL	
	Mean	Sigma	Mean	Sigma	Mean	Sigma
Experience:						
Naive	.503	.092	16.62	1.65	45.42	18.63
Trainee	.599	.075	17.96	1.44	45.54	13.59
Ph.D.	.575	.062	17.57	1.72	45.89	13.78
Voice:						
Client	.588	.145	18.31	3.17	45.67	18.72
Actor	.506	.167	15.97	3.99	43.65	18.71
Script	.584	.313	17.88	2.79	47.53	19.87
Content:						
#1	.623	.151	17.69	3.64	51.57	18.65
#2	.447	.141	15.58	3.35	41.22	18.63
#3	.608	.096	18.88	2.62	44.06	18.74

Because of the relatively large differences between the average scores under the Client and Expert keys, it was considered expedient to test the significance of these differences by the simple nonparametric sign test (13, pp. 430–435). The results of this test clearly verified the observation that the judges within each group show greater agreement with the Experts' predicted responses for the clients than with the responses of the clients themselves.

Discussion

From the results of this study, it appears that judges with one or more years of clinical experience are able to make more accurate predictions about clients' responses to certain personality measures than can be predicted by untrained judges. But among judges who have had at least one year of training, increased experience does not appear to be associated with a significant increase in the accuracy of predictions. There is the problem of whether these results imply that whatever skill is added to the individual's ability to make judgments along these lines is acquired rapidly, during the early period of graduate training, and that increments thereafter are too gradual to make for significant experience level differences. The results could be interpreted also to suggest that selection may be the important factor here: that certain kinds of people, better than average in ability to judge the responses of people who are upset, enter clinical psychology and that training to the level of the Ph.D. does not increase their skills along these lines. Another possibility is that perhaps the clients used in this study placed the more experienced Ph.D. judges at a disadvantage, since they work mostly in hospital settings, and the clients used in the study were outpatients of a university counseling service.

In the present study, unlike the Luft studies where clinicians were no more accurate than nonclinicians, brief interview selections were presented to the judges. It could be that the trainee and the experienced clinician are able to formulate better personality appraisals on the basis of little data than are naive judges. Luft's findings with one-hour interviews that experienced clinicians are no more accurate than inexperienced judges suggests the possibility that as the amount of interview material available as a basis for judgment increases, the superiority of the experienced judges over naive judges may diminish. Or, there may be differences due to age differences, since in the present study, the Naive judges were undergraduates. But these speculations must await empirical testing.

While it is true that, for all three personality measures employed in this study, a discrepant (the actor's) voice lowered the accuracy of clinical judgment, reading the interview selection allowed more accurate predictions of *Q* sorts and Gough ACL responses than did listening to the client. This corresponds to Luft's report that student readers did as well as student listeners in predicting responses to objective questionnaires, and seems related to Borke and Fiske's finding that reading interview statements was associated with as accurate predictions as occurred after face-to-face interaction with the client.

In this study, no significant Voice-by-Experience or Content-by-Experience interactions occurred. The experience groups were more accurate than the Naive judges over all conditions studied, and not just one content or one voice.

Conclusions

1. Two groups of judges with clinical training and experience, the Trainees and the Ph.D. judges, predicted client responses more accurately than the Naive judges.

2. Between the experienced groups, however, there occurred no statistically significant differences in the accuracy of prediction of client responses.

3. A voice discrepant from the original speaker of the interview content lowered the accuracy of the judgments for all groups of judges.

4. Different interview contents appeared to be related to differences in accuracy of judgments about the client, but the instrument being used appeared to be an important variable here.

5. The script condition, with total absence of voice cues, was associated with the most accurate predictions with two of the three measuring instruments used.

6. There was no statistically significant interaction between experience and voice, or between experience and content.

7. The judges, regardless of experience level, were in greater agreement with what a consensus of "experts" predicted about client responses than in predicting the actual responses made by the clients themselves.

Received July 17, 1957.

References

1. Borke, Helene, & Fiske, D. W. Factors influencing the prediction of behavior from a diagnostic interview. *J. consult. Psychol.*, 1957, **21**, 78–80.
2. Cline, V. B. Ability to judge personality assessed with a stress interview and sound film technique. *J. abnorm. soc. Psychol.*, 1955, **50**, 183–187.
3. Gage, N. L. Explorations in the understanding of others. *Educ. psychol. Measmt*, 1953, **13**, 14–26.
4. Gough, H. G. *Reference handbook for the Gough Adjective Check List*. Berkeley, Calif.: Univer. Calif. Inst. Pers. Assessment and Res., 1955.
5. Guilford, J. P., & Martin, H. G. The Guilford-Martin Inventory of Factors GAMIN. Beverly Hills, Calif.: Sheridan Supply, 1943.
6. Hilden, A. H. *Manual for Q-Sort and random sets of personal concepts*. Webster Groves, Mo.: Author, 1954.
7. Kostlan, A. A method for the empirical study of psychodiagnosis. *J. consult. Psychol.*, 1954, **18**, 83–88.
8. Lindquist, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1956.
9. Luft, J. Implicit hypotheses and clinical predictions. *J. abnorm. soc. Psychol.*, 1950, **45**, 756–759.
10. Luft, J. Differences in prediction based on hearing versus reading verbatim clinical interviews. *J. consult. Psychol.*, 1951, **15**, 115–119.
11. Soskin, W. F. Bias in postdiction from projective tests. *J. abnorm. soc. Psychol.*, 1954, **49**, 69–74.
12. Stern, G., Stein, M., & Bloom, B. *Methods in personality assessment*. Glencoe, Ill.: Free Press, 1956.
13. Walker, Helen M., & Lev, J. *Statistical inference*. New York: Holt, 1953.

Journal of Consulting Psychology
Vol. 22, No. 4, 1958

A Note on the Reliability of the Leiter International Performance Scale 1948 Revision[1]

Heber C. Sharp
Utah State University

Since the Leiter International Performance Scale (LIPS) was developed, most of the reported research has been designed to explore the validity of the LIPS by comparing it with other measures of intelligence. Reliability studies seem to have been neglected. Leiter (1) states that two types of reliability data are available. The type reported most frequently compares the standard deviation of the LIPS results with those of the accepted instrument. Leiter reports a split-half reliability coefficient of .89 (1, p. 30). Tate (2), following Thorndike's outline (3, pp. 93–95) of Holt's analysis procedure to estimate reliability, reports a coefficient of .75. The following data are concerned with a test-retest reliability coefficient.

Procedure. The LIPS was administered to 48 retarded school children in September or October and again in April or May with at least a six-month interval between tests. The *S*s ranged in age from 8 years and no months to 16 years and 5 months. 10 *S*s were attending a special school for seriously retarded children, 21 were attending special classes for retarded children, and 17 were attending classes with other children in the regular schools.

Results. Statistical data revealed for the initial test an M of 65.1, SD 13.67, and the range of 38–88. The retest M was 67.77, SD 13.73, and the range 41–95. The reliability coefficient was found to be .91. A t test to determine the significance of the difference between the means yielded a t of 3.15 ($p < .01$).

Discussion. The retest method may introduce a spurious factor due to carry-over from one administration to the other. *S*s frequently mentioned memory for one or more test items. This might account for the difference found between the means. Only 6% of the *S*s had the same score on both test and retest, 27% earned lower scores on the retest, and 67% showed increases from 1 to 14 IQ points on the retest. For this retarded group, some items seemed to be improperly placed in the scale. Further empirical work concerned with this problem is suggested.

The reliability coefficient of .91 indicates both stability and consistency of the LIPS as a measuring instrument for *S*s at the lower end of the intelligence distribution. The similarity of the variance measures supports further the reliability of the LIPS. Since the *S*s were drawn from a restricted range of ability, thus limiting the spread of scores, the high coefficient does not imply necessarily reliability for all types of subjects.

Received August 12, 1957.

References

1. Leiter, R. G. *Part I of the manual for the 1948 revision of the Leiter International Performance Scale.* Chicago: C. H. Stoelting, 1952.
2. Tate, Mariam T. The influence of cultural factors on the Leiter International Performance Scale. *J. abnorm. soc. Psychol.,* 1952, **47**, 497–501.
3. Thorndike, R. L. *Personnel selection test and measurement techniques.* New York: Wiley, 1949.

---

[1] This study was supported in part through a grant from Utah State University research program.

Journal of Consulting Psychology
Vol. 22, No. 5, 1958

The Prediction of Length of Stay in Psychotherapy[1]

Maurice Lorr, Martin M. Katz, and Eli A. Rubinstein
U. S. Veterans Administration

Much needed by mental hygiene clinics is a method of early identification of patients who for various reasons terminate treatment after only a few visits and without the advice or consent of the therapist. Various surveys have shown that anywhere from 30% to 60% of patients referred to a mental hygiene clinic and accepted for psychotherapy will terminate within the first six treatment visits (4, 7). In a substantial number of cases termination occurs even before the initial interview with the therapist. Early identification of premature terminators would permit the clinic to adapt its intake procedure in appropriate ways. Special procedures designed to motivate or orient potential terminators might be instituted. Or, other treatment modalities such as tranquilizer drugs might be utilized or devised.

In an earlier study, four short psychological tests and questionnaires were found to be predictive of length of stay in treatment (9). Patients from 10 Veterans Administration outpatient clinics throughout the country who remained in psychotherapy for at least six months were compared with those who terminated within a month. The patients consisted of nonpsychotic males in intensive individual therapy who had not been seen in treatment previously. The items were selected on the basis of a double cross-validation on two random halves of a sample of 128 cases. Results indicated that remainers may be characterized as less nomadic, less impulsive, less rigid in personal attitudes, and more self-dissatisfied than terminators. In addition, remainers tend to have more education, perhaps because of greater goal-directed persistence, and they tend to come from higher socioeconomic levels than terminators.

One aim of the present study was to test the validity of the predictive test battery on a wider range of clinic cases. Another objective was to test a series of research hypotheses concerning the characteristics that differentiate patients who terminate prematurely from patients who remain in treatment. These hypotheses grew out of a survey of the literature and out of the above mentioned research results of the initial study (2, 6, 12).

The specific hypotheses, in each case from the standpoint of remainer characteristics, are as follows:

1. Terminators are more likely to have a history of frequent trouble with the law, lack of impulse control, hostility to authority, lack of goal persistence, and lack of personal ties or loyalties.
2. Terminators are less self-dissatisfied.
3. Terminators are less likely to report anxiety.
4. Terminators have more limited vocabularies.
5. Terminators are more authoritarian.
6. Terminators' socioeconomic level is lower.

Method

A nation-wide group of 13 clinics agreed to collect data over a period of three months on all cases considered for psychotherapy (group or individual).[2] Patients examined for hospitalization, neurological disorders, or for brief (half hour or less) irregular treatment were not included.

---

[1] From Veterans Administration, Veterans Benefits Office, Washington, D. C.

[2] We are indebted to the VA Mental Hygiene Clinics and respective clinic staffs in the following cities for collaborating in the collection of the basic data: Baltimore, Brooklyn, Chicago, Cleveland, Denver, Miami, Newark, Pittsburgh, St. Louis, San Antonio, San Francisco, Seattle, and Washington, D. C.

The intake interviewer, usually a social worker, completed a form at the close of the interview describing such patient characteristics as his occupation, annual earnings, and presenting problems. In addition, the interviewer estimated how long the patient would remain in treatment if accepted. In this way the test battery's success in predicting length of stay could be compared with the judgment of an interviewer unaided by test information. The therapist filled out a data sheet when a patient discontinued treatment prior to the 26th visit for any reason, or at the end of six months of treatment if the patient was still in treatment at that time. The therapist also indicated the reason for termination if this occurred, the frequency and duration of visits, and the type of treatment given.

The measures used, given below, were self-administerable and required 45 minutes or less for completion.

1. A 39-item Behavior Disturbance scale taken from a longer unpublished inventory devised by Applezweig and Dibner. The true-false items include elements primarily of a biographical nature designed to elicit information concerning the extent of certain behavior in the patient's past. The items ask questions concerning lack of personal ties or loyalties, lack of impulse control, restlessness, frequent trouble with the law, lack of ethical standards, hostility to authority, and lack of goal persistence.
2. A Self-Rating scale consisting of 18 five-point graphic rating scales. First the patient rates his actual self and then he rates himself the way he would like to be on the same scales. The difference between the self-rating and the ideal-rating provides a measure of self-dissatisfaction.
3. A 30-item version of the Taylor Manifest Anxiety Scale which purports to measure manifest anxiety (10).
4. A 15-item multiple choice vocabulary test, which represents a modification of the Thorndike selection of words in the Stanford-Binet Vocabulary Test.
5. A 20-item *F* scale taken from Adorno, Frenkel-Brunswik, et al. (1). The scale purports to be a measure of authoritarianism and conventionalism as defined by the authors. The patient indicates his degree of agreement with each statement on a four-point scale.

In all, approximately 300 usable cases were received from the 13 clinics. Cases were excluded from the study if (*a*) the patient terminated prior to six full months from the date of initiation of treatment with mutual consent of the therapist and was rated improved; (*b*) the interval between the initial intake interview and the initial treatment was over 10 weeks; (*c*) the patient did not receive an average of at least one treatment every four weeks during his stay in treatment; (*d*) treatment was terminated for such reasons as transfer, hospitalization, ineligibility, or relocation to another city; and (*e*) the patient was female. The first four qualifications made the group a more clear-cut sample of terminators and remainers. The sample was confined to male patients because so few female veterans are seen in VA clinics.

The criterion used in this study was number of weeks in treatment. The number of treatments received was considered as a possible criterion but rejected since the number of treatments scheduled varies from patient to patient. Furthermore, the number of treatments is likely to be a less reliable figure in most clinics than sheer length. However, the correlation of .60 between the two variables indicated that the measures have much in common. The entire initial sample of 291 cases was randomly split into two samples: A and B. Patients were so assigned that each clinic and each therapist had approximately equivalent representation in Subsamples A and B. The distribution by weeks of treatment was also equated in the two samples.

The criterion distribution was found to be U-shaped; only 60 cases were in treatment for more than 6 weeks but less than 26 weeks. It was also evident, as in the original study, that the middle group could not be differentiated by the test battery from the two end groups. Accordingly patients with 6 weeks of treatment or less were designated Terminators, and patients with 26 weeks or more of treatment were designated Remainers. The middle group of patients staying from 7 to 26 weeks was not used in the present analysis. Each subsample consisted of 115 cases, with 57 Terminators and 58 Remainers in each sample.

To test the research hypotheses, and to identify the items in Sample A significantly related to the dichotomous criterion, point biserial correlations and phi coefficients were computed and tested for significance.[3] The items within a given inventory or test which correlated with the Terminator–Remainer cri-

[3] The statistical assistance of Elizabeth Turk is gratefully acknowledged.

terion at the .05 level of significance or better were combined with unit weights into a subtest. The answer sheets were then rescored for each subtest, and the subtests were correlated with the criterion. The Lubin-Summerfield square root procedure (8) was applied to the six variables most highly correlated with the criterion. This procedure selects from a group of variables the minimum set of effective variables which will result in the highest possible multiple correlation coefficient with the criterion.

A configural approach was next applied to the minimum set of effective variables selected by the square root procedure. The configural approach was utilized because it takes into account the interaction among variables, and any nonlinear relationships which are not included in a linear multiple regression equation. Each of the four subscores selected was dichotomized at the mean and each patient in Sample A was allocated either to the upper or lower half of each variable. In each case, a high score represented a Remainer score, while a low score typified the Terminator. The combination of Upper and Lower categories on the four variables yields 16 patterns. Each patient was classified into one of the 16 patterns. The percentage of Terminators and Remainers falling into each pattern was then determined. Finally the percentages were transformed into a series of "configural" scores ranging from 0 to 9 by the following equation (11):

$$S = 9\left[\frac{\begin{matrix}\text{\% Remainers in given} \\ \text{pattern}\end{matrix} - \begin{matrix}\text{Lowest \% Remainers} \\ \text{in any pattern}\end{matrix}}{\begin{matrix}\text{Highest \% Remainers} \\ \text{in any pattern}\end{matrix} - \begin{matrix}\text{Lowest \% Remainers} \\ \text{in any pattern}\end{matrix}}\right]$$

These configural scores were then correlated with the dichotomous criterion.

The beta weights derived from Sample A were then used to score the 115 cases in Sample B plus an additional 16 cases received late from one clinic. Similarly, the 16 patterns derived from Sample A were used to score the cases in Sample B. Point-biserial correlations were then computed with length of stay to determine the degree of cross-validity obtaining for the four subtests selected.

To complete the double cross-validation, the items for Sample B were independently correlated with the criterion as had been done for Sample A. An item was selected for the final set if the item-criterion correlation was significant at the .05 level or better in both samples (a one-tailed test).

The Findings

The first five research hypotheses were tested through a comparison of Terminator and Remainer scores on the a priori keys for the five tests. A result was regarded as significant if the point-biserial correlation was significant at the .05 level or better on both Samples A and B. None of the a priori keys proved to be differentiating in both samples although all scores were consistently in the predicted direction.

Differences found between Terminators and Remainers in annual earnings, occupational level, and highest grade completed were also in the predicted direction. However, the differences were not at the required significance levels in both samples. An examination of the data by clinic suggested that there might be substantial clinic differences. Consequently, data on annual earnings from five of the clinics with the largest samples were subjected to a two-way analysis of variance. No significant clinic differences and no significant interactions between clinics and the criterion were found. However, an F test significant at the .001 level and a point biserial of .28 indicated that annual earnings were indeed predictive of termination. However, an analysis of occupational level and grade level by somewhat different procedures failed to reveal significant differences between Terminators and Remainers for the same five clinics.

The item analysis of Sample A identified six differentiating items in the F scale, eight in the Behavior Disturbance scale, eight items in the Manifest Anxiety scale, and five items in the Self-Rating scale. Each set of items was combined into a subscale. The correlations among the four subscales, the Vocabulary Test, the Social Worker's Prediction, and the criterion were subjected to the square root analysis previously described. The criterion correlations of the six variables analyzed by this procedure are presented in Table 1. A multiple correlation of .67 significant at the .01 level was obtained between the F subscale, the Behavior Disturbance subscale, the Mani-

Table 1

Correlations of Subscales and Social Worker's Prediction with Terminator–Remainer Criterion in Samples A and B

Variable	Correlation Sample A	 Sample B
F scale	.32**	.07
Manifest Anxiety scale	.30**	.25**
Behavior Disturbance scale	.39**	.31**
Self-Rating scale discrepancy	.32**	.10
Vocabulary test	.12	.21*
Social Worker Prediction	.36*	.36*

** Significant at .01 level.
* Significant at .05 level.

fest Anxiety subscale, the Social Worker's Prediction, and the criterion. Application of Sample A beta weights to Sample B scores resulted in a combined test-score-criterion correlation of .39, which is also significant at the .01 level.

The configural approach to the minimum set of four variables yielded a correlation of .62 for Sample A. Next, Sample B was scored on Sample A patterns. The correlation of Sample B patients' configural scores with the criterion yielded a correlation coefficient of .43. The configural approach thus results in as effective a separation of Terminators and Remainers on cross-validation as multiple regression.

To complete the cross-validation, the individual items significantly correlated with the criterion in Sample A were checked on Sample B and those significant in Sample B were checked on A. By this procedure, 22 items proved to be significantly differentiating at the .05 level or better (one-tailed test) and in the same direction in both samples. The entire sample of 230 cases was then rescored on the significant 21 items allocated to three subtests (Anxiety scale, F scale, Behavior Disturbance scale). Configural scores based on these three subtests and the social worker's prediction correlated .50 with the criterion.

The actuarial items were also correlated with the criterion in both Samples A and B. Only race and religion proved to be consistent in predicting termination of treatment. Negro patients were more likely to be terminators than other patients (r_{pb}'s of .25 and .18). Jewish patients compared to all others were more likely to be remainers (r_{pb}'s of .17 and .17). Tests were also made for type of psychiatric disorder, current psychiatric disability rating, number of previous courses of treatment, a listing of ten major problem areas (e.g., problems with authority figures, vocational adjustment, marital conflict), and marital status. None were found to relate to the criterion at significant levels in both samples. Scheduled treatment frequency, length of treatment interview, type of treatment (group, individual, or both) showed no significant relation to the criterion. Likewise the analysis indicated that the sex, discipline, and length of experience of the therapist were unrelated to the criterion.

The possibility remained that a position response set was operative in several of the original scales and was in part responsible for the scale-criterion correlations. The index of response set was defined as $R_T/N_T - R_F/N_F$, where R_T is the number of items keyed True and marked True by the examinee, and R_F is the number of items keyed False and marked False by the examinee. N_T and N_F are the numbers of items keyed True and keyed False respectively. Correlations of response-set scores with the criterion were computed separately for the Manifest Anxiety and Behavior Disturbance scales. The correlations obtained were .08 and −.05 respectively. Evidently, then, response sets are not responsible for these scale-criterion correlations. However, the question remained whether the response set was acting as a suppressor variable. To check this possibility, content and set scores were correlated in each scale. Neither coefficient of −.14 nor .03 was significant. Thus response set as here defined exerted no influence on the scale-criterion correlations.

Discussion

As we indicated earlier, all of the scores based on the a priori keys were in the hypothesized direction on both samples but not at acceptable levels of significance. Of the 89 items in the Behavior Disturbance scale, the F scale, and the Anxiety scale, 22 items were significantly correlated with the criterion at the .05 level or better in both samples. Only one of these significant items was keyed in a direction not hypothesized. Consequently no other items are inconsistent with the hypotheses. Do the remaining 21 items tend to be consistent with and to support the research hypotheses?

The Behavior Disturbance items will be considered first. The research hypothesis states that Terminators are more likely than Remainers to report a history of antisocial behavior, to lack impulse control, and to lack loyalties or personal ties. The following responses seem to support these conjectures:

> I have been arrested three times or more. (True)
>
> I have never been in a reform school, prison, workhouse, or jail. (False)
>
> "Every man for himself" is the wisest rule to follow. (True)
>
> I lose interest in things which I cannot get or do right away. (True)
>
> I have often spent more money than I had by borrowing on the spur of the moment. (True)
>
> Right now I have some money saved up. (False)
>
> I sometimes break a date with someone without telling him about it. (True)
>
> When we go out together, I sometimes walk off and leave my friends without telling them about it. (True)

The third hypothesis states that Terminators are less likely to report anxiety than Remainers. The Taylor scale items correlated significantly with the criterion were the following:

> I am often sick to my stomach. (False)
>
> I wish I could be as happy as others. (False)
>
> I find it hard to keep my mind on a task or job. (False)
>
> Life is often a strain for me. (False)
>
> I am not very confident of myself. (False)
>
> My hands and feet are usually warm enough. (True)

Denial of these reactions certainly lends support to the postulated lack of, or failure to report, anxiety in the Terminator patient.

Are Terminators more "authoritarian" than Remainers as hypothesized? Terminators tend to agree with the following F scale items which are significantly related to the criterion:

> When a person has a problem or worry, it is best for him not to think about it, but to keep busy with more cheerful things.
>
> Nowadays when so many different kinds of people move around and mix together so much, a person has to protect himself especially carefully against catching an infection or disease from them.
>
> Sex crimes, such as rape and attacks on children, deserve more than mere imprisonment; such criminals ought to be publicly whipped, or worse.
>
> People can be divided into two distinct classes: the weak and the strong.
>
> There is hardly anything lower than a person who does not feel a great love, gratitude, and respect for his parents.
>
> Someday it will probably be shown that astrology can explain a lot of things.
>
> Nowadays more and more people are prying into matters that should remain personal and private.

The statements describe an individual who believes in the repression of conflict, who is lacking in psychological sophistication and insight, and who identifies with authority. To the extent that these statements represent measures of "authoritarianism," the fifth research hypothesis is supported.

We next turn to a consideration of the patterns established. Let the order of the subscales be as follows: (*a*) F scale, (*b*) Manifest Anxiety scale, (*c*) Behavior Disturbance scale, (*d*) Social Worker Prediction. Let each frequency distribution be dichotomized at the mean. Next let U represent the upper or Remainer half and L represent the lower or Terminator half of each distribution. Then a patient allocated to the Terminator pattern LLLL is low on all four variables and is characterized as authoritarian, nonanxious, and as reporting an antisocial history; he is also rated Terminator by the social worker. An example of a Remainer pattern is UULU which represents a patient who is nonauthoritarian, anxious, with a history of antisocial behavior, but rated Remainer by the social worker.

An examination of the five most discriminating Remainer patterns is revealing. Any combination of three or more Upper scores is predictive of remaining in treatment. The Manifest Anxiety scale seems to be most crucial as judged by the percentage of successes; Remainers are anxious. The Terminators also are best identified by any combination of three or more Lower scores. Terminators may also be identified successfully by any three Low scores or by any two Low scores providing one is on the Anxiety scale.

The discriminatory validity of the configural approach may be judged by the data presented in Table 2. The optimum cutting score for separating Terminators and Remainers in the cross-validation sample was

Table 2

Number of Terminators (T) and Remainers (R) Classified into Each of the 16 Patterns (Patients in both Samples A and B are scored on Sample A patterns)

	Sample A		Sample B	
Pattern	T	R	T	R
LLLL	9	0	6	0
LLLU	2	1	3	1
LLUL	3	1	3	0
LLUU	2	1	6	5
LULL	7	5	8	4
LULU	5	5	8	7
LUUL	4	2	0	3
LUUU	0	3	1	4
ULLL	4	0	9	0
ULLU	4	1	3	1
ULUL	6	4	6	2
ULUU	3	12	7	12
UULL	3	3	4	2
UULU	0	3	2	7
UUUL	1	4	0	1
UUUU	0	11	4	7

first estimated. It is the point of intersection of the frequency distributions of configural scores for the two criterion categories (4). With this cutting score, 56 (or 80%) of the Terminators are correctly classified by the test battery. This compares favorably to the 55% Terminators one would estimate from the base rate in the present population. In all, 90 (or 71%) of Terminators and Remainers are correctly classified.

The Remainer is thus seen as an anxious, self-dissatisfied individual with some psychological insight who is willing to explore his personal problems with others. He has some sense of loyalty to others and tends to persist in activities he undertakes. He is not likely to have been involved in antisocial acts. On the other hand, the Terminator either is not anxious or does not admit to being anxious and self-dissatisfied. He is likely to have had a history of antisocial acts, he admits to being undependable and impulsive, and may be authoritarian or rigid in his social attitudes.

Agreement of these findings with previously reported studies is substantial. Frank *et al.* (3), in their summary, list the following attributes as related to remaining in therapy: social class, education and occupation, fluctuating illness with manifest anxiety, readiness to communicate distress and personal liabilities, influenceability, social integrity, and perseverance. Similarly, Hiler (5) lists education, intelligence, level of ambition, tendency to be introspective, degree of psychological sophistication, degree of anxiety experienced, and felt dissatisfaction as important variables. Other reported studies on the role of social class place a greater importance on socioeconomic status than was evidenced in the present study.

Undoubtedly premature termination is also a function of the therapist and the type of therapy offered. The therapist–patient relation and the competence of the therapist may both relate to length of stay in treatment. It would also seem probable that in order to reduce the rate of dropout, therapeutic procedures must be developed for handling the relatively uneducated, poorly motivated, and psychologically unsophisticated patient. Perhaps special orientation procedures are needed. Possibly drugs should be used as adjunctive treatment. In any case, premature termination in large numbers suggests that present clinic practice does not meet the needs of a considerable proportion of patients referred for treatment.

Summary

The aims of the study were to check the validity of a predictive test battery and to test six research hypotheses concerning characteristics that differentiate patients who terminate prematurely from those who remain

Table 3

Number of Terminators (T) and Remainers (R) Correctly Classified in Cross-Validation Sample B when Cutting Score is Optimum

Test classification	Actual classification		
	T	R	Total
T	56	22	78
R	14	34	48
Total	70	56	126

for six months or longer. Data on five tests were collected on 300 cases seen in 13 clinics. Scores based on the a priori keys did not differentiate the groups at an acceptable level of significance although all were in the expected direction. However, the item analysis resulted in three subscales that did tend to support four of the hypotheses. The best three subscales and the social worker's prediction combined configurally correlated .43 with the Terminator–Remainer criterion in the cross-validation sample. Compared to Terminators, Remainers tend to be more anxious, more self-dissatisfied, and more willing to explore personal problems with others. Remainers also are less likely to have a history of antisocial acts and are more dependable, more controlled, and more persistent in tasks undertaken than Terminators.

Received November 1, 1957.

References

1. Adorno, T. W., Frenkel-Brunswik, Else, Levinson, D. J., & Sanford, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. Auld, F., Jr., & Myers, J. K. Contributions to a theory for selecting psychotherapy patients. *J. clin. Psychol.,* 1954, **10**, 56–60.
3. Frank, J. D., Gliedman, L. H., Imber, A. D., Nash, E. H., Jr., & Stone, A. R. Why patients leave psychotherapy. *Arch. Neurol. Psychiat.,* 1957, **77**, 283–299.
4. Guilford, J. P. *Fundamental statistics in psychology and education*. New York: McGraw-Hill, 1950. P. 371.
5. Hiler, E. W. Psychological factors associated with premature termination of psychotherapy. Unpublished doctoral dissertation, Univer. of Michigan, 1954.
6. Imber, A. D., Nash, E. H., Jr., & Stone, A. R. Social class and duration of psychotherapy. *J. clin. Psychol.,* 1955, **11**, 281–284.
7. Kurland, S. H. Length of treatment in a mental hygiene clinic. *Psychiat. Quart. Suppl.,* 1956, **30**, 1–8.
8. Lubin, A., & Summerfield, A. A square root method of selecting a minimum set of variables in multiple regression: II. A worked example. *Psychometrika,* 1951, **16**, 425–437.
9. Rubinstein, E. A., & Lorr, M. A comparison of terminators and remainers in outpatient psychotherapy. *J. clin. Psychol.,* 1956, **12**, 345–349.
10. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.,* 1953, **48**, 285–290.
11. Wherry, R. J. Maximal weighting of qualitative data. *Psychometrika,* 1944, **9**, 263–266.
12. Winder, A. E., & Hersko, M. The effect of social class on the length and type of psychotherapy in a Veterans Administration Mental Hygiene Clinic. *J. clin. Psychol.,* 1955, **11**, 77–79.

Journal of Consulting Psychology
Vol. 22, No. 5, 1958

Suggestibility and Dependency[1]

Marvin Zuckerman and Hanus J. Grosz
Institute of Psychiatric Research, Indiana University Medical Center, Indianapolis

In the book *Explorations in Personality* (4), a study relating hypnotizability to personality needs is reported. Hypnotizability was found to correlate .43 with need for Deference and − .44 with need for Autonomy. The Sway Test has been suggested as a predictor of hypnotizability (3) and has been found to load highly on a factor of "primary suggestibility" (2).

The Sway Test was used as part of two other studies going on at this institute. The *S*s were student nurses. Of the *S*s who had taken the Sway Test, 27 had also taken the Edwards Personal Preference Schedule (EPPS) (1), a test of manifest needs based on the Murray system; and 19 *S*s had also taken the Thematic Apperception Test (TAT) which had been scored by Murray's need system. The authors decided to compare relevant personality needs of high and low groups on the Sway Test on the assumption that suggestibility is related to dependency traits. Sway tendency was measured in inches of positive sway on two trials. The group was divided at the median into high and low swayers.

Although a number of need scores were examined in comparing the two groups, three are directly relevant because of the White study: Deference, Autonomy, and Succorance. Succorance and Deference seem to measure "Dependency" while Autonomy would define the other pole of this dimension. The low swayers scored significantly higher on Autonomy on the EPPS than the high swayers (p below .01). The other differences were not significant. The Autonomy score on the EPPS indicates endorsement of statements which express a desire ". . . to be independent of others in making decisions, to feel free to do what one wants . . . to avoid situations where one is expected to conform . . ." (1, p. 5).

The high swayers scored significantly higher on Succorance attributed to the hero in TAT stories (p below .05) and higher on Succorance attributed to all characters in stories (p below .05). The other differences were not significant. A need for Succorance on the TAT indicates the number of instances where the *S* saw the character in the story asking for or receiving help, sympathy, or support from another individual or when the character is disturbed over loss of a source of love and support.

The findings from the two personality tests are congruent and indicate that a person who is suggestible is likely to be a person with strong dependency needs, while a person who resists suggestion is more liable to have stronger needs for independence or autonomy.

Brief Report.
Received March 19, 1958.

References

1. Edwards, A. L. *Edwards Personal Preference Schedule, Manual.* New York: Psychological Corp., 1954.
2. Eysenck, H. J., & Furneaux, W. D. Primary and secondary suggestibility: An experimental and statistical study. *J. exp. Psychol.,* 1945, **35,** 485–503.
3. Furneaux, W. D. The prediction of susceptibility to hypnosis. *J. Pers.,* 1946, **14,** 281–294.
4. Murray, H. A. *Explorations in personality.* New York: Oxford Univer. Press, 1938.

---

[1] An extended report of this study may be obtained without charge from Marvin Zuckerman, Institute of Psychiatric Research, 1100 West Michigan Street, Indianapolis 7, Indiana, or for a fee from the American Documentation Institute. Order Document No. 5581, remitting $1.25 for photocopies and $1.25 for microfilm.

Journal of Consulting Psychology
Vol. 22, No. 5, 1958

Success and Failure in Client-Centered Therapy as a Function of Initial In-Therapy Behavior[1]

William L. Kirtner and Desmond S. Cartwright
University of Chicago

In a previous paper (4), the authors reported that certain differences in client personality structure at the outset of therapy appear to be significantly related to differences of length and outcome in client-centered therapy. The present study investigates the hypothesis that the manner in which clients initially present their problems and their initial in-therapy mode of approach to the resolution of those problems are also significantly related to differences in length and outcome.

Methods and Procedure

The Instrument

On the basis of the preceding hypothesis, five descriptions of initial in-therapy client behavior were formulated, and each designated by one of the letters, V, W, X, Y, or Z. A complete statement of each of these descriptions follows:

V: Immediately deals with a feeling-in-relationship problem and has already somewhat localized a rather specific source or area of difficulty. Deals very much with what he says and does, how he acts and feels in situations, and discusses the interpersonal effects of these. He is quite internally focused and has a very strong and very apparent drive to generate and examine impulses, thoughts, ideas, despite resultant fear, guilt, sadness, etc. The person orients himself in the therapy situation as if saying: "This is my response to such-and-such and this is the kind of situation in which I find myself; now assuming that I somehow contribute to this situation, I want to alter my responses and behaviors and resolve or diminish the disturbances I feel."

W: Immediately deals with feeling-in-relationship problems but has not clearly differentiated a specific source of difficulty. That is, relational difficulties are perceived, but the feelings involved are not understood; no clear connection with the situation is apparent. There is strong drive toward clearer differentiation and understanding of where the source of interpersonal or self-disturbance lies and how it occurs. There is as well a general concern to discover just how and where in his disorders he himself is contributing, and effort is spent in driving himself to look at generated impulses for his contributions to his undesirable life situations so that he might change and resolve his disturbances.

X: Vacillates between dealing with relationship problems and discussing externals and listing attributes of situations and of others. He may openly exhibit emotional behavior: e.g., depression, crying, fear, anger, etc. He may give way to a mood or feeling experience of the moment but usually expresses or discusses these feelings in terms of external cause rather than in terms of "this is my response to such and such" (external); though, to repeat, there is usually a vacillation between expression of mood or emotion and discussion of it in terms of internal and external causes. In a word, he vacillates between expressing himself about basic feeling and interaction problems and the manifestations of those problems in very general terms without definite pointing to his own contributions to situations and his responsibility in them; and yet shows definite indications now and then of feeling that he is in some way a contributor and has some responsibility for the situations.

Y: Does not deal with feeling-in-relationship problems but discusses external manifestations of internal difficulties; or discusses feelings as if they are external objective things to be intellectually named, labeled, or categorized (e.g., "I use a lot of attention-getting mechanisms in social situations"). There is a rigorous intellectual control of impulses so that they are shunted into structured categories and explanatory generalities which seem to give a measure of momentary satisfaction and comfort. Such dealing, shunting, and applying of generalities appear to

---

[1] This investigation is based upon a dissertation submitted by the senior author to the University of Chicago in partial fulfillment of the requirements for the degree of doctor of philosophy. It was supported in part by research grants from the National Institute of Mental Health, National Institutes of Health, U. S. Public Health Service, and from the Ford Foundation.

Table 1

Results of Application of the Five Descriptions to First Therapy Interviews of Twenty-four Cases for Which Both Length of Therapy and Therapist Success Rating Were Known to One of the Judges

Client	Failure zone (13–21 interviews)	Short failure (12 or less interviews)	Short success (12 or less interviews)	Long success (More than 21 interviews)	Long failure (More than 21 interviews)
1	Z	Y	V	W	X
2	X	Y	V	W	..
3	Y	Y	V	X	..
4	Y	Z	..	W	..
5	Y	Z	..	..	..
6	Y	X	..	..	..
7	Y	Z	..	..	..
8	..	Z	..	..	..
9	..	Z	..	..	..

represent problem-resolving activity and are usually followed by a similar procedure about another behavioral area. There is strong avoidance of discussion on the feeling level in relationships, but listing of attributes of feeling life may be done intellectually and analytically.

Z: Deals with problems as though they are almost entirely external to him. The localization of source of difficulties is seen to be vaguely "beyond" the person, the focus being quite external. His approach is almost that of listing attributes of people and situations, with very little self-responsibility understood. It is as if at times the person is attempting to resolve his difficulties by explaining the people with whom he is involved, disturbed about, guilty toward, etc. It is sometimes as if he is saying: "Things should be different, and if they were, I'd be all right." He often appears to be describing the various facets of disturbing situations and relationships as if they are entirely outside him and as if then waiting for something to be done about them. Since he focuses the problems outward, he seems to be asking: "What can be done to change this so that I won't feel disturbed?" The therapy situation thus seems to be tripolar: the client, the problems, the counselor. There is avoidance of discussion of internal feelings in relationships, even though feeling may be apparent in voice tone, gesture, words used, etc.

Subjects

The total sample for this study is composed of 42 cases, all seen by therapists at the Counseling Center, The University of Chicago, during the period 1949–54. They were participants in various research programs being carried out at the Counseling Center at that time. The sample was composed of 21 students, 21 nonstudents. There were 20 females and 22 males. The mean age was 27.9 years; the age range was from 19 to 41 years. Forty-one of the cases had male therapists; in one case the therapist was female.

Twenty-four cases from the total sample represent only the extremes on a therapist 9-point success rating scale—that is, ratings of 1 through 4 being failures, and 7 through 9 being successes. The remaining 18 cases of the total sample represent the more middle range, points 5 and 6, of the 9-point rating scale, as well as its extremes.

Procedure

The instrument described in a previous section of the present paper was applied to the first therapy interview only of each of the 42 cases constituting the sample. Both length of therapy and therapist success rating were known by the senior author (who acted as one of the judges) for 24 of the 42 cases. Of these 24 cases, 14 were examined and rated jointly by him (Judge I) and another judge (II). The remaining 10 of these 24 cases were rated independently by the two judges.

For the remaining 18 cases of the total sample, neither judge knew length of therapy or therapist success rating for any case.[2] Application of the descriptions to 14 of these 18 initial therapy interviews was made independently by each judge. Judge I alone applied the instrument to the remaining 4 cases.

---

[2] D. S. Cartwright collected the interview material and coded it for use by the two judges.

Results

Joint rating of 14 of the 24 cases for which the senior author alone knew both length of therapy and therapist success rating resulted in the two judges reaching immediate concensus. Independent application of the instrument to the remaining 10 of these 24 cases resulted in perfect agreement between judges.

Table 1 presents the results of application of the instrument to these 24 cases grouped according to length-by-outcome.

The foregoing applications indicate the following predominant relationship between client groups and instrument descriptions: (*a*) the failure-zone group may be mainly characterized as Y, with two exceptions which are Z and X; (*b*) the short failures may be characterized as Z and Y, with one exception which is an X; (*c*) the short successes may all be characterized as V; (*d*) the long successes may be characterized as W, with one exception which is an X; and (*e*) the single long failure is described as X.

Since the block of 18 cases represented both the middle range and the extremes of the success–failure scale, it was necessary to redefine prediction points for success and failure with respect to these cases. Cartwright (2) had demonstrated the mean success rating for the population of clients terminating therapy at the Counseling Center, The University of Chicago, to be approximately 5.00 on the therapist 9-point success rating scale. Thus, for the purposes of classifying these 18 cases, the writers chose to define therapeutic success as a rating of 6 or more on that scale, and failure as a rating of 5 or less.

Table 2 presents the results obtained from the judges' application of the descriptions to the first therapy interviews in the block of 18 cases, within the framework of the actual length-by-outcome groupings.

Inspection of the data shown in Table 2 indicates that the two judges agreed on 13 of 14 ratings with respect to application of description to client, a result significant at better than the .005 level of confidence.

The results also indicate the following predominant relationships between descriptions and clients: (*a*) Short failures may be characterized as Z or Y, with one exception which is an X; (*b*) short successes were not predicted, though two are given such actual ratings by the therapist, being rated by the two judges as short failures, Z; (*c*) long successes may be characterized as W or X, with one exception which was rated by both judges as a short failure, Z. Comparison of these findings with those in the previous block of 24 cases indicates that, with a few important exceptions, the predominant relationships between instrument descriptions and length-by-outcome groupings appear to hold.

Something should be said concerning the important exceptions mentioned in the preceding paragraph. First, with respect to the

Table 2

Results of Application of the Five Descriptions to First Therapy Interviews in Eighteen Cases, within the Framework of Their Actual Length-by-Outcome Groupings

	Short failure (12 or less interviews)		Short success (12 or less interviews)		Long success (More than 21 interviews)	
Client	Judge I	Judge II	Judge I	Judge II	Judge I	Judge II
1	X	X	Z	Z	W	Not Rated
2	Y	Not Rated	Z	Z	W	W
3	Y	Y			X	Not Rated
4	Y	Y			X	X
5	Z	Not Rated			X	X
6	Z	X			X	X
7	Z	Z			Z	Z
8	Z	Z				
9	Z	Z				

two cases rated by the therapists as short successes and by the judges as short failures, perhaps little more can be said than that the criteria used by the therapists in these cases must have been markedly different from those employed by the judges. A reauditing[3] of the initial interview in these two cases results in the same conclusion as that given in the first place—namely, that within the framework of the instrument developed for this study, no other prediction than short failure appears possible with respect to these two cases.

It should be remembered in this connection that the instrument predicted the length of therapy quite adequately in both cases. With respect to length of therapy, no cases fell within the failure zone (13–21 interviews); only two cases fell within the short success category, and there were no long failure cases in this subsample. In order to demonstrate statistically something of the predictive value of the instrument with regard to length of therapy, the subsample of 18 cases was dichotomized on the length dimension. Since Cartwright (2) had also demonstrated that approximately 50% of the clients entering therapy at the Counseling Center leave therapy at or before the 15th interview and approximately 50% remain in therapy for 16 or more interviews, these divisions were chosen to form the dichotomy. A two-by-two table formed on these two divisions of length shows that only one case was incorrectly placed; the result is significant at better than the .002 level of confidence. A comparable table formed on the dichotomy of therapeutic success (ratings of 6 or more on the 9-point scale) and failure (ratings of 5 or less on the 9-point scale) shows that only three cases were incorrectly placed; the result is significant at better than the .015 level of confidence (3).

Of considerable importance is the distribution of the X descriptions throughout the sample. A number of cases were designated as X, and each length-by-outcome group contains one or several of those so designated. Since the instrument did in general differentiate correctly the length-by-outcome group into which each case should be placed, the evidence is clear that the X description is not adequately defined for a specific and particular task. Such a conclusion derives from the observation that cases rated as "long therapy X" are almost invariably rated by the therapist as successful, whereas those rated "short therapy X" are almost invariably rated by the therapist as failures. The basis for this distinction is not clear. Perhaps further investigation of these sorts of matters will lead to descriptive behavioral differences on the basis of which clearer and more reliable predictions can be made. No further resolution of rating discrepancies apparent in this study can be made at the present time.

Discussion

Following examination of factors such as age, sex, status (student-nonstudent), and therapist, in relation to length and outcome of therapy, Cartwright concluded that (1, p. 362) ". . . certain individual differences between clients . . . give rise to different kinds of therapeutic process." In addition, he hypothesized that the failure-zone may be characterized by (1, p. 363) "a drastic behavioral manifestation of resistance," suggesting that "If such a view were confirmed by further research, it would raise seriously the question of whether or not client-centered therapists should modify their approach to therapy to include the possibility of temporarily directive behavior at a point during this critical failure zone when it seems likely that a client will leave."

The present study supports Cartwright's hypothesis that individual differences between clients account, at least in part, for different kinds of therapeutic process. However, as a result of the present study, it would seem that earlier rather than later in-therapy modification of therapist approach might be desirable or required if failure-zone as well as short failure cases are to remain in therapy. Whether or not such modification of therapist approach should be toward "directive" operations remains unclarified.

In contrast to the earlier hypothesis concerning resistance during the critical failure-zone period, the present study suggests that the resistance may be identifiable at the outset of therapy, though such a suggestion does not rule out the possibility of increasing resistance and drastic display of it in later periods of therapy.

---

[3] Perhaps it should be stated that an original auditing and fitting of a description requires approximately two and one-half to three hours. Acoustical equipment now in common use would probably reduce the time considerably.

It seems reasonable to suppose that the manner in which the client conceptualizes and attempts to resolve his problems will have much to do with whether or not he achieves resolution. In client-centered therapy as practiced today, there is no place for activity on the part of the therapist specifically directed toward helping the client to change an ineffective mode of approach to the resolution of his problems. Persistent understanding and acceptance of a "resistant" mode of approach may well do little more than reinforce that approach.

Whether modes of approach to problems are indeed modifiable at all is, of course, an open question waiting upon controlled investigation. But the writers conclude that in subsequent study of the therapeutic process, special attention must be paid to that aspect of interaction between client and therapist; namely, to the client's approach to his problems and the therapist's responses to that approach.

Summary

The first therapy interviews of 42 clients seen by client-centered therapists were examined through the application of an instrument describing five differing modes of client initial approach to the resolution of personal-life difficulties.

Results from two independent judges showed that the descriptions differentiated rather clearly the length-by-outcome groups constituting the total sample of the study. A discussion of these results followed, and it was concluded that the manner in which clients approach and conceptualize their problems will have much to do with whether or not resolution is achieved.

Received July 31, 1957.

References

1. Cartwright, D. S. Success in psychotherapy as a function of certain actuarial variables. *J. consult. Psychol.*, 1955, **19**, 357–363.
2. Cartwright, D. S. Synthesis of process and outcome research. Discussion Paper, Counseling Center, Univer. of Chicago, Vol. II, 1956.
3. Finney, D. J. The Fisher-Yates Test of Significance in 2 x 2 contingency tables. *Biometrika*, **35**, Parts I and II, May, 1948. Pp. 145–156.
4. Kirtner, W. L., & Cartwright, D. S. Success and failure in client-centered therapy, as a function of client personality variables. *J. consult. Psychol.*, 1958, **22**, 259–264.

Journal of Consulting Psychology
Vol. 22, No. 5, 1958

Some Psychological Correlates of Field Independence[1]

David Marlowe
Ohio State University

Recent research has dealt with the relationship between personality factors and field independence. Witkin et al. have reported that "the personality characteristic most closely related to field dependence or independence seems to be the tendency toward active coping with or passive submission to the environment" (2, p. 474). Field independent persons were described as more analytical and more capable of mastering environmental forces than field dependent persons who were characterized as passive and submissive toward authority. Wertheim and Mednick (1) reported a significant relationship between field independence and need achievement. The present study attempted to test the above findings using objective paper-and-pencil tests. It was hypothesized that field independence is positively correlated with needs Achievement, Autonomy, Dominance, and Intraception and negatively correlated with Succorance.

Sixty-nine undergraduate *S*s (57 female, 12 male) were administered the Edwards Personal Preference Schedule (EPPS) and the Thurstone adaption of the Gottschaldt Embedded Figures Test (EFT). On the EFT, *S* was shown 18 simple geometric designs, each followed by four complex designs with instructions to find the simple design in the complex designs. *S*'s score on the EFT was the number correct in the 20 minutes alloted for the task.

Product-moment correlations were obtained between scores on the EFT and the scores on each of the 15 needs measured by the EPPS. *S*s' scores on the EFT ranged from 9 to 34 with a mean of 21.3. The greater the *S*'s score on the EFT, the greater the field independence.

Of the five needs hypothesized to be relevant, only two yielded significant correlations with field independence: Intraception, .34 ($p < .01$) and Succorance, $-.30$ ($p < .02$). It can be concluded that in the present sample, field independence is associated with the need to be analytical in regard to the behavior and motives of one's self and others (Intraception), and with a relative absence of passive-dependent needs (Succorance). These results offer only partial support for Witkin's findings. Most noteworthy is the failure of Autonomy and Dominance to yield significant correlations. These needs may be considered similar to Witkin's "active coping" and "mastery of environmental forces" factors. The failure to obtain results similar to Wertheim and Mednick's for need Achievement may be due to differences between the studies in measuring instruments. In particular, n Ach as measured by fantasy materials may not be equivalent to n Ach as measured by the EPPS.

Brief Report.
Received April 28, 1958.

References

1. Wertheim, J., & Mednick, S. A. The achievement motive and field independence. *J. consult. Psychol.*, 1958, 22, 38.
2. Witkin, H. A., Lewis, H. B., Hertzmon, M., Machover, K., Meissner, P. B., & Wapner, S. *Personality through perception*. New York: Harper, 1954.

---

[1] An extended report of this study may be obtained without charge from David Marlowe, Ohio State University, Room 304, Arps Hall, Columbus 10, Ohio, or for a fee from the American Documentation Institute. Order Document No. 5692, remitting $1.25 for microfilm or $1.25 for photocopies.

Journal of Consulting Psychology
Vol. 22, No. 5, 1958

Psychoanalysts as Expert Observers of the Therapy Process[1]

Richard L. Cutler, Edward S. Bordin, Joan Williams, and David Rigler[2]
University of Michigan

Our research group has as its major interest the study of the process of psychotherapy, that is, the analysis of the interaction between patient and therapist (2, 3, 6). One immediate goal is the testing of several specific hypotheses concerning the relationship of the therapist's interpretive activity to the manifestation of resistance in the patient. Due to the relative recency with which interest in this area of therapy research has developed, however, we have found it necessary to digress in order to solve some of the measurement and methodological problems involved in operationalizing our variables (1, 5). This paper reports a further research concerned with the effects of varying conditions of rating and classes of raters upon one of our key variables, Depth of Interpretation.

In a previous study clinical psychologists made ratings of therapy interviews on Depth of Interpretation under varying conditions of presentation, context, and unit size. In that study, as in the present one, interpretation was defined as "any behavior on the part of the therapist that is an expression of his view of the patient's emotions and motivations . . ." (4, p. 247). The greater the disparity between the view expressed by the therapist, and the patient's awareness of these emotions and motivations, the deeper the interpretation. A seven-point scale for rating Depth of Interpretation was developed by the Method of Equal Appearing Intervals from 70 statements descriptive of therapist behavior. An analysis of variance revealed: (*a*) Raters were able to apply this particular scale so as to distinguish between interviews; (*b*) a single, over-all rating of an interview led to a deeper rating than that obtained from the mean of response-by-response ratings; (*c*) ratings made from typescripts did not differ from those based upon tape recordings.

All of these findings supported our expectations. Contrary to what we anticipated, however, there were no differences ascribable to varying amounts of context. That is, it made no difference whether the raters were given the immediately preceding interview in its entirety before rating the crucial interview, or whether they made their ratings using only the therapist's responses, with all patient material deleted.

It was primarily because we were concerned about this inability of our raters to utilize increasing amounts of contextual information to increase their interjudge agreement that the present study was designed. We first sought logically to identify those elements in the stimulus material which were potential contributors to the variance in rating Depth of Interpretation. The first of these hypothetical contributors, Element A, may be thought of as consisting of valid cues from the stimulus material which provide evidence as to the level of the patient's awareness of his emotions and motivations, as to the therapist's expression of *his* view of the patient's emotions and motivations, and, hence, as to the disparity between them. Thus, Element A represents the "true" variance of Depth of Interpretation. The second, Element B, consists of a set of cues which lead to reliable judgments based upon some implicit, absolute conception of what constitutes a deep or shallow interpretation. The third, Element C, consists of irrelevant or interfering cues from either the patient's or the therapist's comments, which serve to mislead or confuse the raters. A fourth element, which is relatively independent of the stimulus material, but which interacts with Elements A, B, and C in the total rating process, is the sensitivity to the various proportions of A, B, and C in the stimulus material which is available in the particular judge population.

---

[1] This study was carried out under the auspices of United States Public Health Service Grant M-516, "Analyses of Therapeutic Interaction," E. S. Bordin and R. L. Cutler, Principal Investigators.

[2] Now at Children's Hospital, Philadelphia, Pennsylvania.

It is immediately apparent that Element A should be present to a greater degree in a situation where stimulus material is presented with more, rather than less, context. Element B may or may not be increased by increasing the amount of contextual material available to the judges during the rating process. Element C is almost certainly increased, since the judges are asked to maintain an awareness of considerable additional material, and must, in addition, decide whether a given bit of stimulus material is valid and relevant (Element A) or irrelevant (Element C).

We were thus led to consider the possibility that the judge population was crucial in this particular rating process, and that a particular kind of sensitivity to the nuances of therapeutic interaction was necessary to enable judges to capitalize upon the additional contextual material to increase their interjudge reliability. Given adequate training in the perception of cues in the patient's and therapist's responses relevant to the rating of Depth of Interpretation, and in the disregarding of irrelevant or confusing cues, our judges should demonstrate more convincing interjudge agreement than had been found in the earlier study.

Table 1

Assignment of Raters to Experimental Conditions[a]

Typescript (TY)		Recording (RE)	
Therapist Only (TO)	Preceding Interview (PI)	Therapist Only (TO)	Preceding Interview (PI)
[b]A, I, 4, 1	A, I, 2, 2	A, I, 3, 4	A, I, 1, 3
A, II, 4, 1	A, II, 2, 2	A, II, 3, 4	A, II, 1, 3
A, III, 2, 4	A, III, 4, 3	A, III, 1, 1	A, III, 3, 2
F, IV, 2, 4	F, IV, 4, 3	F, IV, 1, 1	F, IV, 3, 2
F, V, 3, 3	F, V, 1, 4	F, V, 4, 2	F, V, 2, 1
F, VI, 3, 3	F, VI, 1, 4	F, VI, 4, 2	F, VI, 2, 1
A, VII, 1, 2	A, VII, 3, 1	A, VII, 2, 3	A, VII, 4, 4
F, VIII, 1, 2	F, VIII, 3, 1	F, VIII, 2, 3	F, VIII, 4, 4

Note.—The latin square was originally designed to pair one analyst and one fledgling in each of the four order–condition–case sequences. However, due to difficulties in obtaining qualified raters, it was necessary to modify the design as above.

[a] This table may be recomposed into a modified latin-square design.

[b] Key to notation. Capital letters refer to Analyst (A) or Fledgling (F). Roman numerals indicate judges' identifying numbers. First arabic numeral in sequence indicates interview number. Second arabic numeral in sequence indicates order in which the particular judge performed this segment of his rating task.

Procedures

Accordingly, we enlisted as expert judges in this study four psychoanalysts and four psychiatrists who had completed their personal analyses and were undergoing control training in psychoanalysis, but who had not yet been admitted to full standing in either the local or national analytic societies. The former group are hereafter called analysts; the latter, for want of a better term, fledglings.

Our rationale for the selection of this particular judge population was as follows: even though our clinical psychologist judges had had experience as psychotherapists, only a few had completed a personal analysis, and none had had the benefit of psychoanalytic training. We felt that the emphasis in analytic training upon the development of an acute self-awareness, with its consequent increased perceptual and interpersonal sensitivity and objectivity, might better enable the analysts to judge the degree of the patient's awareness of his emotions and motivations, to assess the therapist's statements concerning his view of the patient's emotions and motivations, and hence more reliably to rate Depth of Interpretation. These judges should also be better equipped to weed out irrelevant cues, and perhaps to be less led astray by invalid cues of Element B. In short, we hoped that the analysts and fledglings would be better able to capitalize upon increased amounts of contextual information to increase their interjudge agreement by making maximal use of Element A and by limiting the distracting effects of Element C.

In addition, by including the fledglings, we hope to assess the effects of varying amounts

Table 2

Breakdown of Sums of Squares and F Tests for Analysis of Variance

Source of variation	Sum of squares		df		Mean square		F	
Sequences	4.26		3		1.420		33.02**	
Raters within sequences	5.46		4		1.365		31.74**	
Total between raters	9.72		7					
Order	.05		3		.017		—	
Interviews	1.93		3		.643		14.95**	
Experimental conditions	.28		3		.093		2.16	
Presentation		.08		1		.080		1.86
Context		.01		1		.010		—
Presentation × Context		.19		1		.190		4.42
Pooled error	.65		15		.043			
Total within raters	2.91		24					
Grand total	12.63		31					

** Significant at the .01 level.

of training and experience within the group which had had the benefit of formal analytic training.

A modified latin-square design was projected which was essentially a replication of the design of our earlier study. It permitted the systematic variation or control of two levels of context (therapist responses only vs. preceding interview) and two methods of presentation (typescript vs. tape recording). It also enabled us to assess the effects of practice, order of presentation of the interviews, judges, and differences due to the interviews themselves. In addition, it was possible to obtain a rough estimate of the effect of experience within this judge population (analysts vs. fledglings) by means of a t test, although this difference is confounded with order and conditions. In the actual process of rating, each judge rated each of the four interviews response-by-response on the seven-point Depth of Interpretation rating scale (Table 1).

Results

The results of the analysis of variance are presented in Table 2. A word should be said at this point about this analysis. We were concerned with the effect of the varying conditions of rating upon the means of the response-by-response ratings in an interview. These means are extremely stable, and thus the total sum of squares is extremely small. In order to be conservative, we included the second and third order interactions in the error term.

The analysis of variance showed that, as in the previous study, this second judge population was able to apply the Depth of Interpretation rating scale to differentiate between interviews. By far the greatest amount of the total variance was contributed by the judges; neither order of presentation nor varying amounts of context contributed significantly to the variance. However, the context by presentation interaction approached the .05 level of significance.

A second portion of our data analysis was directed toward the interjudge agreement or rating reliability across and within the varying conditions. A summary of the interjudge reliabilities by cases is presented in Table 3. The product-moment correlations are remarkably constant across Cases I, II, and III, but in Case IV there is a significant drop in interjudge reliability.[3] A possible explanation for this difference lies in the fact that while Cases

---

[3] Significance of difference between ranges of correlations was assessed in every case by the Median test, a nonparametric statistic which does not require the assumption of underlying normality in the population of correlations (7).

I, II, and III are counseling or psychotherapy sessions with normal or neurotic subjects, Case IV is from a series of interviews between a schizophrenic patient under insulin treatment and an inexperienced psychotherapist.

In Table 4 are presented the interjudge reliability coefficients by conditions of presentation and context. Significantly higher interjudge agreement is found when the analyst-fledgling judges rate under identical conditions of presentation and context than when ratings are made under totally different conditions. However, the rather startling result is revealed that when this group of judges had additional contextual material at its disposal, the interjudge reliabilities were significantly lower than when they made their ratings on the basis of information from the therapist's responses alone. When the results for the analysts and fledglings are viewed separately, the number of correlations becomes so small that statistical significance is not obtained, but the trend is equally apparent in both the analysts and the fledglings and cannot be considered to be the result of the difference in experience. In addition, agreement was slightly, although not significantly, better when the ratings were based upon typescripts rather than upon tape recordings.

Data from the earlier study with the psychologists were further analyzed to permit the assessment of the effects of varying cases and conditions of presentation and context upon their ratings. The results of these analyses are presented in Tables 5 and 6. Once again, Case IV is seen to yield consistently lower interjudge agreement, although the Median test reveals only the difference between Cases III and IV to be significant. There is no significant difference in interjudge agreement which can be attributed to different methods of presentation, and while ratings made under the preceding interview context condition yield slightly higher interjudge agreement than those made under the "therapist only" condition, the difference is not significant. Additionally, identical conditions of presentation and context produce no higher agreements among the psychologists than do totally different conditions.

Table 3

Interjudge Reliabilities by Cases
(Analyst-Fledgling group)

Case	Range	Median
I	−.03 to .56	.34
II	−.05 to .65	.31
III	.09 to .53	.32
IV	−.15 to .48	.16

Note.—Cases I, II, and III show significantly ($p < .05$) higher interjudge agreement than Case IV. Cases I, II, and III do not differ significantly.

Table 4

Interjudge Reliabilities by Conditions of Presentation and Context

(Analyst-Fledgling group)

Conditions	Range	Median
[a]Identical	.01 to .56	.37
[b]Totally different	−.05 to .58	.26
[c]Identical context (TO)	.05 to .65	.32
[d]Identical context (PI)	.00 to .46	.25
[e]Identical presentation (RE)	−.15 to .53	.25
[f]Identical presentation (TY)	.03 to .53	.31

Note.—Identical vs. totally different—$p < .05$.
Therapist Only context vs. Preceding Interview context—$p < .01$.

[a] Includes only those correlations between raters operating under the same conditions of context and presentation.

[b] Includes only those correlations between raters operating under both different context and presentation conditions.

[c] Includes those correlations between raters operating under the Therapist Only context condition, irrespective of method of presentation.

[d] Includes those correlations between raters operating under the Preceding Interview context condition, irrespective of method of presentation.

[e] Includes those correlations between raters rating recordings, irrespective of context condition.

[f] Includes those correlations between raters rating typescripts, irrespective of context conditions.

Finally, comparisons were made of the interjudge reliabilities within and among the psychologist, analyst, and fledgling groups. In general, we find higher interjudge agreement among the psychologists than among either the analysts or fledglings. When ratings are made under identical conditions of presentation and context, the psychologists yield a range of interjudge reliabilities from .04 to .58, with a median of .42, while the comparable figures for the analyst group are − .01 to .56, median .22; and for the fledglings, .09 to .25, median .14. These differences among the groups, due to the small numbers of cases involved, are not significant.

A comparison of the interjudge reliabilities within each of the three groups' ratings made under entirely different conditions of presentation and context reveals a range of .07 to .62 for the psychologists, with a median of .42; range, .35 to .42, median .36 for the analysts; range −.02 to .43, median .31 for the fledglings. Once again, these differences fail to reach statistical significance, probably because of the small number of correlations involved.

The amount of agreement among the three groups was determined by a series of cross correlations. That is, the ratings of the analyst group were correlated with the ratings of the psychologist group, the analysts with the fledgings, etc. These results may be summarized by saying that in general, the psychologists agreed more with both the analysts and fledgling groups than the latter agreed with each other.

Discussion

This study supports our earlier evidence that raters are able to apply the Depth of Interpretation scale so as to differentiate between interviews, and indicates further that levels of experience and exposure to formal analytic training do not substantially modify this ability. There is consistent evidence that the method of presentation of the interview material (typescript vs. recording) has no effect either upon mean level of rating or interjudge agreement. While the differences in level of interjudge agreement among the three groups are small, there is a consistent tendency for the psychologist group to agree among themselves to a slightly greater extent than either the analysts or fledglings. In addition, we find that our hypothesis concerning the effect of the increased sensitivity resultant from analytic training upon the perception of cues provided by additional contextual information is not confirmed. Instead, we find the paradoxical result that among the analyst-fledgling group, additional context results in a significant decline in interjudge agreement.

Table 5

Interjudge Reliabilities by Cases

(Psychologist group)

Case	Range	Median
I	−.04 to .62	.37
II	.13 to .63	.37
III	.31 to .64	.48
IV	−.05 to .73	.26

Note.—Significant difference between III and IV, $p < .01$.

Table 6

Interjudge Reliabilities by Conditions of Presentation and Context

(Psychologist group)

Conditions	Range	Median
[a]Identical	.04 to .58	.42
[b]Totally different	.07 to .62	.42
[c]Identical context (TO)	.09 to .63	.35
[d]Identical context (PI)	−.05 to .64	.43
[e]Identical presentation (RE)	.04 to .59	.41
[f]Identical presentation (PI)	.03 to .73	.37

Note.—None of the three comparisons reveals a significant difference.

[a] Includes only those correlations between raters operating under the same conditions of context and presentation.

[b] Includes only those correlations between raters operating under both different context and presentation conditions.

[c] Includes those correlations between raters operating under the Therapist Only context condition, irrespective of method of presentation.

[d] Includes those correlations between raters operating under the Preceding Interview context condition, irrespective of method of presentation.

[e] Includes those correlations between raters rating recordings, irrespective of context condition.

[f] Includes those correlations between raters rating typescripts, irrespective of context conditions.

It may be possible to account for the generally higher interjudge agreement among psychologists in terms of their greater familiarity with rating tasks of the sort demanded by these studies. This does not, however, explain why the analyst group was handicapped by the additional context. It is possible that when this group was faced only with the therapist responses, it was able to impose a generally more meaningful rationale upon the data than when the additional context provided a much larger number of alternative hypotheses. Perhaps with the fuller context of having reviewed *all* of the preceding interviews, the analyst group would tend to converge on a more unanimous judgment. The possibility that analytic training leads to a somewhat greater ability to entertain multiple alternative hypotheses in the face of larger amounts of contextual information should not be overlooked.

Whatever may account for the few differences between psychoanalysts and clinical psychologists in rating Depth of Interpretation, it seems very clear that the two groups are essentially interchangeable as observers in studies which use this method of rating. Even though some may question the validity of this particular method of differentiating Depth of Interpretation, it is fairly representative of the commonly used methods. Therefore, one may proceed with the confidence that choice within this particular range of observers will not have profound influences on the results obtained.

We continue to be troubled by the generally low reliability of the Depth of Interpretation scale, although we have previously demonstrated the possibility of increasing interrater agreement by means of specific instruction in the application of the scale, and by using ratings of somewhat larger segments of the interview. We continue to feel that the approach to psychotherapy research by means of response-by-response analysis is the most promising of several alternatives, and are convinced that the variable Depth of Interpretation is a meaningful and useful one in psychotherapy research. A recent study by Speisman (8) lends support to this contention.

Summary

In order to explore the possibility that increased perceptual sensitivity to the subtleties of the therapy relationship would allow raters to increase their interjudge agreement in rating Depth of Interpretation, four psychoanalysts and four analysts in training were enlisted as expert judges. These judges rated four therapy interviews under different conditions of presentation and context.

An analysis of variance carried out on a modified latin-square design revealed that these raters could distinguish between interviews, but that neither the method of presentation of the interviews nor the amount of contextual information available to the judges had any systematic effect upon the means of response-by-response ratings.

A correlation analysis revealed that this group of raters had generally lower interjudge reliabilities than a group of psychologists who had previously rated the same interviews under the same conditions. The analyst-fledgling group among themselves showed significantly better interjudge agreement when their ratings were made under conditions of minimal context (therapist only condition) than when increasing amounts of context were available (preceding interview condition).

These results were discussed in terms of the greater familiarity of psychologists with rating tasks of the sort used, and the possibility was raised that analytic training made it possible to entertain a greater number of alternative hypotheses (and thus to decrease rating reliability) in the face of added contextual information. The essential identity of the two populations of judges was pointed out, and it was suggested that one may proceed with the confidence that choice within this particular range of observers will not have profound influences on the results obtained in studies in which this and similar rating scales are used.

The implications of the study for the continued application of the Depth of Interpretation rating scale to the process analysis of psychotherapy were discussed.

Received September 9, 1957.

References

1. Bordin, E. S., Cutler, R. L., Dittmann, A. T., Harway, N. I., Raush, H. L., & Rigler, D. Measurement problems in process research on psychotherapy. *J. consult. Psychol.,* 1954, **18,** 79–82.
2. Cutler, R. L. The relationship between the therapist's personality and certain aspects of psychotherapy. Unpublished doctoral dissertation, Univer. of Michigan, 1953.
3. Dibner, A. S. The relationship between ambiguity and anxiety in a clinical interview. Unpublished doctoral dissertation, Univer. of Michigan, 1953.
4. Harway, N. I., Dittmann, A. T., Raush, H. L., Bordin, E. S., & Rigler, D. The measurement of depth of interpretation. *J. consult. Psychol.,* 1955, **19,** 247–253.
5. Raush, H. L., Sperber, Z., Rigler, D., Williams, Joan V., Harway, N. I., Bordin, E. S., Dittmann, A. T., & Hays, W. A dimensional analysis of depth of interpretation. *J. consult. Psychol.,* 1956, **20,** 43–48.
6. Rigler, D. Some determinants of therapist behavior. Unpublished doctoral dissertation, Univer. of Michigan, 1956.
7. Siegel, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.
8. Speisman, J. C. The relationship between depth of interpretation and verbal expressions of resistance in psychotherapy. Unpublished doctoral dissertation, Univer. of Michigan, 1956.

Journal of Consulting Psychology
Vol. 22, No. 5, 1958

An Analysis of Patient–Therapist Compatibility

E. Wesley Hiler[1]
University of Michigan

One of the problems which confronts most mental hygiene clinics is premature termination of psychotherapy. Many patients go through all the admission procedures, begin a course of long-term therapy, but discontinue treatment after only a few sessions.

Several articles have dealt with the possible causes of premature termination of therapy (2, 7, 9, 10, 18) and a number of attempts have been made to predict which patients will terminate (3, 12, 14, 15, 17, 22). Most of the prediction studies have been concerned with the personality of patients, as inferred from test records and biographical data, that might be related to their unwillingness to remain in treatment. Very few studies have dealt with characteristics of therapists which might be responsible in part for patients discontinuing therapy.

The present study is an attempt to discover whether different types of therapists tend to lose or hold in treatment different types of patients. It is not concerned with differences in the drop-out rate for different therapists, but rather the *type* of patient most apt to continue or discontinue treatment with various types of therapists. It is thus a study of patient-therapist compatibility. It is therefore in line with the contemporary emphasis on field theory, which conceives of therapy as an interpersonal process in which the patient's behavior and growth are seen not as a function of the patient's personality alone, nor of the therapist's personality and skill alone, nor the mere sum of these, but is rather dependent on the particular nature of the interaction between the patient's characteristics and the therapist's characteristics. The results of such studies could eventually be put to use for the purpose of selecting the therapist most suitable for each patient. At present, patients are assigned to or selected by therapists at random or else on an intuitive basis rather than on the basis of empirical research.

In studies reported previously (14, 15) dealing with the prediction of premature termination of psychotherapy by means of the Rorschach test, it was found that the total number of responses given turned out to be the best predictor. Three other studies have reported somewhat similar findings (3, 12, 17). In a number of clinics, therefore, it appears that patients who drop out of treatment within a few sessions usually give fewer responses on the Rorschach than patients remaining in treatment for a longer period of time.

A high response total on the Rorschach has been attributed to or found related to a number of personality and attitudinal variables including: productivity (5, 24), achievement (26), intelligence (3, 20), verbal fluency (24), ego-involvement in the task (8), cooperativeness (3), responsivity (6), ease at forming associations (1), conflict awareness (25), high energy level (6), and initiative (24). Low response total has been attributed to excessive use of repression (16, 25), depression (5), incapacitating anxiety (11), passivity (24), a tendency to give up easily (24), and a hysterical indifference towards symptoms (24).

The response total on the Rorschach seems thus to be related to a number of intellectual, temperamental, and motivational variables, all reflecting a type of productivity or drive which one would expect to be required for active participation in psychotherapy. In fact some studies have reported statistically significant relationships between the number of Rorschach responses given by the patient and ratings of improvement in psychotherapy (21, 23).

Although for the clinic as a whole, patients who were unproductive on the Rorschach tended to drop out of therapy prematurely, it was noticed by the author that for a few therapists this general finding did not seem to apply. Whether or not productivity on the Rorschach differentiated between patients remaining and patients terminating seemed to depend on which therapist was doing the treatment. A series of studies was therefore carried out to determine whether different types of therapists tended to hold in treatment different types of patients and, specifically, whether some therapists are able to hold in treatment the unproductive or unmotivated patients who usually do not remain long in treatment, and whether some therapists lose many of the productive and motivated patients who usually remain in treatment.

---

[1] This article is based on a doctoral dissertation submitted to the Department of Psychology, University of Michigan, 1953. The author is now at Agnews State Hospital.

Subjects and Setting

The present study was carried out in the VA Mental Hygiene Clinic at Detroit, Michigan. The subjects are therefore veterans (almost all males) who applied for or were referred for psychotherapy at this clinic. All such patients are entitled to free psychiatric treatment because of a service connected disability. Many are referred for treatment after a routine pension examination and often have the erroneous impression that their pension will be jeopardized if they do not report for treatment, and some have misgivings about the fact their pensions will be taken away if they recover as a result of treatment.

A wide variety of diagnostic categories are represented among the patients at this clinic, the most common being anxiety reaction, psychoneurosis, latent schizophrenia, and character disorder. Almost all the patients are from urban areas. Most are employed as skilled or unskilled factory labor; some hold clerical or sales positions; only a few are in the professions or are business executives. The socioeconomic level of the patients is predominately lower and lower-middle class. About half the patients had not completed high school. Upon admission to the clinic the patients are interviewed and usually are given the Rorschach test and one or two other psychological tests.

The therapists at this clinic are for the most part analytically oriented and tend to emphasize long-term methods of therapy. Thus most patients remaining less than 20 sessions are considered to have terminated before completion of the treatment.

Procedure

The records of patients seen by certain therapists were taken from the files and classified with respect to duration of therapy. As in the previous study (14), only extreme cases were used. Patients dropping out of treatment within five sessions without extenuating circumstances or the therapist's recommendation were considered to have terminated prematurely. Those remaining at least 20 sessions met our criterion for remaining in treatment. In the Detroit VA Clinic, about 40% of the patients terminate within five sessions and about 30% remain in treatment 20 or more sessions.

The patients were then further characterized in terms of their productivity on the Rorschach test. Patients giving 25 or more Rorschach responses were classified as "productive," and patients giving less than 25 Rorschach responses were classified as "unproductive."

A series of six statistical analyses was then carried out to determine whether the amount of difference in Rorschach productivity between the remainers and terminators is related to the following therapist variables: (*a*) unspecified differences between individual therapists, (*b*) the professional training of the therapist, (*c*) the sex of the therapist, (*d*) the warmth of the therapist, (*e*) the competence of the therapist, and (*f*) the passivity of the therapist.

In order to characterize the therapists along these last three dimensions, ratings were obtained on all the therapists whose patients were used in this study. Ratings were made by the three staff psychologists who were best acquainted with a large number of therapists at the clinic. The ratings were made on a three-point scale in which each rater characterized each therapist as being average, above average, or below average on three different variables: warmth, competence at analytically oriented therapy, and passivity. The raters were instructed to assign an approximately equal number of therapists to each category on the rating scale. The statistical analyses used patients of therapists falling near the extremes of these continua as reflected by agreement by two out of three of the raters that the therapist was either above or below average, the third rater not being in direct contradiction.

The following specific hypotheses related to these rated characteristics of the therapists were formulated and tested:

1. Therapists who appear warm and friendly are able to hold in treatment more of the unproductive or less motivated patients than therapists who appear somewhat cold and distant. Being with a warm and friendly person may be a source of comfort to unproductive and, hence, probably inadequate individuals, and may serve to motivate some of the patients who were not very motivated to undergo psychotherapy at the time they were tested.
2. Therapists who are highly competent at analytically oriented therapy lose fewer well-motivated, productive patients than therapists who are less competent. Although a competent therapist may lose many unmotivated or unproductive patients, because such patients may not be suitable candidates for analytically oriented therapy, he should not lose many productive and well-motivated patients. Therefore, loss of such patients was held to be due to errors in therapeutic technique, which are likely to be made more often by the less competent therapists.
3. Extremely passive therapists hold in treatment fewer of the unproductive patients than therapists employing a more active procedure. It was felt that the unproductive patients included many who were passive dependents and needed a relatively active therapist to lean on or else they would become discouraged and give up.

Results and Discussion

Unspecified Differences Between Individual Therapists

In order to determine whether individual therapists differ in regard to the type of patients they are able to hold in treatment, six therapists were selected who had been at the clinic a long time. A search of the files discovered, for each therapist, seven patients who remained in treatment at least 20 sessions and seven patients who terminated within five sessions. The number of Rorschach responses of each patient was noted and a nonparametric analysis of interaction was carried out in the manner recommended by Mood (19). The analysis permits eliminating any variability in patients' productivity associated with (*a*) the therapists' selection of patients, as would be reflected in differences between the median number of responses of different therapists' patients in general, and (*b*) over-all differences in scores between remainers and terminators in general; thus leaving the sole source of variation the patient-therapist interaction component.

Table 1

Median Number of Rorschach Responses Given by the Remainers and Terminators of Different Therapists

Patients	Therapists A	B	C	D	E	F
Terminators	22	18	21	16	21	15
Remainers	19	35	32	34	18	33

A significant amount of patient-therapist interaction was found ($p = .01$), which implies that individual therapists do differ in regard to the type of patients tending to continue or discontinue treatment.

Table 1 shows the median number of Rorschach responses of the remainers and terminators of each of the six therapists. It will be noted that for two of the therapists there was no appreciable difference in Rorschach productivity between the remainers and terminators. For the other four therapists, however, the patients who remained in treatment gave about twice as many responses as those terminating. Thus remainers and terminators differ in Rorschach productivity for certain therapists but not for others. Some therapists are able to keep the unproductive patients in treatment, other therapists are not. It follows, therefore, that whether or not Rorschach productivity can be used to predict continuance in therapy depends to some extent on the type of therapist involved. The over-all difference in Rorschach productivity between the remainers and terminators for this particular clinic, as reported previously (14, 15), implies that most of the therapists in this clinic are like Therapists B, C, D, and F, rather than like A and E. Other clinics made up of therapists like A and E would probably find no significant difference in Rorschach produc-

Bureau Edn. Psy. Research
DAVID HA... TRAINING COLLEGE
Dated ....
Acc. No ....

tivity between patients remaining in treatment and patients terminating prematurely. These findings may account for some of the inconsistencies in the reports from different clinics concerning their attempts to predict premature termination of therapy from the Rorschach (3, 4, 12, 15, 17, 22).

Professional Training of the Therapist

In order to determine whether the professional training of the therapist makes any difference in regard to the type of patients most apt to continue or discontinue treatment, 78 patients (39 remainers and 39 terminators) were categorized according to the professional training of their therapist (i.e., psychiatrist, clinical psychologist, or psychiatric social worker) and a nonparametric analysis of interaction was carried out. The results of the analysis failed to reveal a significant amount of patient–therapist interaction. Apparently, whether a therapist is a psychiatrist, clinical psychologist, or psychiatric social worker makes little or no difference with respect to the Rorschach productivity of patients tending to remain in treatment.

Sex of the Therapist

In order to determine whether the sex of the therapist makes a difference in regard to the type of patient most apt to continue or discontinue treatment, a nonparametric analysis of interaction was carried out in which patients were categorized according to the sex of their therapist. For this purpose, 40 patients seen by seven female psychologists and social workers were compared with 40 patients seen by seven male psychologists and social workers. Patients seen by psychiatrists were not used since there were no female psychiatrists at the clinic.

Table 2

Median Number of Rorschach Responses Given by the Remainers and Terminators of Therapists Categorized with Respect to Sex

Patients	Therapists	
	Male	Female
Terminators ($N=40$)	15.5	25.0
Remainers ($N=40$)	30.5	22.5

The result of the analysis of interaction was significant at the .05 level of confidence, suggesting that the sex of the therapist is a factor determining whether a given type of patient will remain in treatment or not. Table 2 indicates the median number of Rorschach responses given by the patients remaining and the patients terminating treatment with male therapists and female therapists.

It will be noted that productivity on the Rorschach is not related to continuance in therapy with the female therapists employed at this clinic, whereas it is associated with continuance for patients seen by male therapists.

Of course it would be unwise to generalize this finding to female therapists in general since the records of patients seen by only seven female therapists, all employed in a single clinic, were used in this study. However, it is clear that in the Detroit VA Clinic, at least, the female therapists were able to keep in treatment many of the patients giving few Rorschach responses, i.e., the type of patients who tend to terminate when assigned to male therapists. But, it must also be noted that the female therapists also tend to lose proportionately more of the productive patients than do the male therapists.

The reason is not clear, but certain possibilities may be suggested. The group giving few Rorschach responses might include many passive dependent or frightened individuals who crave a warm supportive relationship with a mother substitute and may feel threatened by father figures. Another possibility is that some patients not really adequately motivated for therapy might relish the opportunity to enjoy a psychologically close relationship with someone of the opposite sex and thus achieve partial gratification of erotic impulses. Still another possibility is that the female therapists might happen to be warmer and more friendly persons than the male therapists at this clinic. It would be interesting to find out whether the same relationships would hold true in other clinics.

Table 3

Rorschach Productivity of Patients Remaining in Therapy with Therapists Categorized with Respect to Warmth

Kind of therapist	No. of patients seen at least 20 sessions by each type of therapist	Proportion who gave less than 25 Rorschach responses
Warm	32	63%
Cold	38	24%

Note.—$t = 3.55$, $p = .0005$. A one-tailed test was used to compute the significance level since the direction of the difference had been predicted.

Personality and Ability of the Therapist

A series of three statistical comparisons was carried out designed to test the hypotheses presented in the procedures section.

Since it was hypothesized that the warm and friendly therapists hold in treatment more of the unproductive patients than do the cold and distant therapists, the proportion of unproductive (R less than 25) patients among those remaining in treatment with the therapists rated most warm was compared with the proportion of unproductive patients among those remaining in treatment with the therapists rated least warm.

The results shown in Table 3 confirm the first hypothesis. Of those patients who remained in therapy with cold therapists, 76% gave productive Rorschachs and only 24% were not productive, whereas of the patients who remained in treatment with warm therapists, 63% were not productive. Thus the cold therapists tended to keep in treatment only the productive well-motivated patients, whereas the warm therapists were able to keep in treatment a sizeable proportion of the unproductive or less motivated patients as well. In fact, the proportion of unproductive patients who remained in treatment with the warm therapists corresponds closely to the proportion of unproductive patients in the total intake population at this clinic. Therefore, if the patient is assigned to a warm and friendly therapist, productivity on the Rorschach can not be used to predict whether or not he will remain in treatment.

The fact that the warm therapists are able to hold in treatment the unproductive patients as well as the productive ones may or may not be desirable. One would have to know something about the outcome of the treatment in order to tell. Perhaps some patients who are really unable to benefit from analytically oriented psychotherapy will remain in treatment because they enjoy the warm and friendly behavior of the therapist toward them. The treatment of other patients, who could derive some real benefit from therapy, might thus be delayed or prevented, since each therapist can only see a limited number of patients.

Since it was hypothesized that the most competent therapists lose fewer productive well-motivated patients than the least competent therapists, the proportion of productive patients among the terminators seen by the most competent therapists was compared with the proportion of productive patients among the terminators seen by the least competent therapists. The findings summarized in Table 4 confirm the second hypothesis. Therapists rated as most competent at analytically oriented therapy do lose a significantly ($p = .03$) smaller percentage of productive well-motivated patients than therapists rated as least competent.

The third hypothesis was not confirmed by the findings. Rated passivity of the therapist was not found to be significantly related to whether or not unproductive patients would remain in treatment. Either the theory which led to this prediction was in error or else the ratings of passivity did not really reflect the therapist's behavior in therapy.

Table 4

Rorschach Productivity of Patients Breaking Off Therapy with Therapists Categorized with Respect to Competence

Kind of therapist	No. of patients breaking off with each type of therapist	Proportion who gave 25 or more Rorschach responses
Most competent	38	29%
Least competent	20	55%

Note.—$t = 1.94$, $p = .03$. A one-tailed test was used to compute the significance level since the direction of the difference had been predicted.

Because of the exploratory nature of this study, only a few of the many possible analyses were made and only the most obvious hypotheses were tested. However, the method used appears to be a useful one and may warrant further consideration. It is suggested that patients be characterized on a number of variables other than mere productivity. For instance, social class, intelligence, anxiety, hostility, depression, introversion, passive dependency and its reaction formation, homosexuality, paranoid tendencies, schizoid tendencies, psychopathy, over- or undercontrol of affect, etc. might be important variables on which patients could be characterized. Likewise a number of variables of possible importance in carrying out effective therapy with different types of patients might be used to characterize the therapists. Hypotheses could then be formulated regarding the best type of therapist for each type of patient. As to the criterion of compatibility, it may be more useful in the long run to use improvement rather than duration of therapy, although the former criterion presents more difficult measurement problems.

Summary

A study was carried out at the Detroit VA Mental Hygiene Clinic to determine whether therapists differ with respect to the type of patients tending to remain in treatment or break off prematurely, and, if so, what characteristics of the therapists may be responsible for the differences in their patients' reactions to them.

In this study, patients were characterized in terms of their productivity on the Rorschach, since this variable had been found previously to be related to premature termination of therapy. Therapists were characterized in terms of professional training, sex, warmth, competence at analytically oriented therapy, and passivity. A series of statistical analyses were then carried out to determine whether these therapist traits were related to the Rorschach productivity of patients continuing or discontinuing treatment.

The results were as follows:

1. Therapists in general differ in regard to the type of patients who continue or discontinue treatment with them.

2. Whether the therapist is a psychiatrist, clinical psychologist, or psychiatric social worker seems unrelated to the type of patients who continue or discontinue treatment.

3. Whether the therapist was male or female did make a difference. The female therapists tended to keep in treatment more of the unproductive patients but also tended to lose slightly more of the productive patients than did the male therapists.

4. Therapists rated as most warm and friendly were able to keep in treatment a larger percentage of unproductive patients than therapists rated as least warm and friendly.

5. Therapists rated as most competent at analytically oriented therapy tended to lose fewer productive patients than therapists rated as least competent.

6. Rated passivity of the therapist seemed unrelated to the productivity of patients remaining in therapy.

It was suggested that the methods employed in this study could be applied to further studies of patient–therapist compatibility, using a greater variety of patient and therapist variables and using improvement rather than mere continuation in therapy as a criterion of compatibility. Such a study would permit a more effective basis for the assignment of patients to therapists than is now in practice. And this would probably decrease the amount of premature termination of therapy and thus add to the clinic's efficiency.

Received October 4, 1957.

References

1. Adcock, C. J. A factorial approach to Rorschach interpretation. *J. gen. Psychol.*, 1951, **44**, 261–272.
2. Adler, M. H., & Gates, P. H. Veteran status as a complication of psychotherapy. *A.M.A. Arch. Neurol. Psychiat.*, 1951, **66**, 652.
3. Auld, F., & Eron, L. The uses of Rorschach scores to predict whether patients will continue psychotherapy. *J. consult. Psychol.*, 1953, **17**, 104–109.
4. Barron, F. Some test correlates of response to psychotherapy. *J. consult. Psychol.*, 1953, **17**, 235–241.
5. Beck, S. J. *Rorschach's test. Vol. II. A variety of personality pictures.* New York: Grune & Stratton, 1945.

6. Beck, S. J. The Rorschach test: A multidimensional test of personality. In H. H. Anderson and Gladys A. Anderson (Eds.), *An introduction to projective techniques*. New York: Prentice-Hall, 1951. Pp. 101–122.
7. Brody, C. Helping a client move into psychiatric treatment through a counseling process. *Jewish soc. serv. Quart.*, 1951, **27**, 265–277.
8. Calden, G. The relationship of varied test definitions and degrees of ego-involvement to Rorschach test performance. Unpublished doctoral dissertation, Univer. of Michigan, 1951.
9. Coble, J. C., Conner, L. E., & Lockwood, W. V. An analysis of the response to follow-up letters to patients who fail to keep mental hygiene clinic appointments. *Amer. J. Psychother.*, 1953, **7**, 289–302.
10. Dengrove, E., & Kutash, S. B. Why patients discontinue treatment in a mental hygiene clinic. *Amer. J. Psychother.*, 1950, **4**, 457–472.
11. Eichler, R. M. Experimental stress and alleged Rorschach indices of anxiety. *J. abnorm. soc. Psychol.*, 1951, **46**, 344–355.
12. Gallagher, J. J. A. comparison of clients who continue with clients who discontinue client centered therapy. In *Group report on a program of research in psychotherapy*. Pennsylvania State Coll., Psychotherapy Research Group, 1953. Pp. 21–38.
13. Gallagher, J. J. Test indicators for therapy prognosis. *J. consult. Psychol.*, 1954, **18**, 409–413.
14. Gibby, R. G., Stotsky, B. A., Miller, D. R., & Hiler, E. W. Prediction of duration of therapy from the Rorschach test. *J. consult. Psychol.*, 1953, **17**, 348–354.
15. Gibby, R. G., Stotsky, B. A., Hiler, E. W., & Miller, D. R. Rorschach criteria for predicting duration of therapy. *J. consult. Psychol.*, 1954, **18**, 185–191.
16. Klopfer, B., & Kelley, D. M. *The Rorschach technique*. Yonkers, N. Y.: World Book, 1942.
17. Kotkov, B., & Meadow, A. Rorschach criteria for predicting continuation in individual psychotherapy. *J. consult. Psychol.*, 1953, **17**, 16–20.
18. Meyer, B. C. Obstacles encountered in recommending psychotherapy: A follow-up study of 400 cases. *J. Mt. Sinai Hosp., N. Y.*, 1948, **15**, 90–96.
19. Mood, I. M. *Introduction to the theory of statistics*. New York: McGraw Hill, 1950.
20. Neff, W. S., & Lidz, T. Rorschach patterns of normal subjects of graded intelligence. *J. proj. Tech.*, 1951, **15**, 45–57.
21. Pollens, B. The relationship between psychological data and prognosis in psychotherapy. Unpublished doctoral dissertation, New York Univer., 1950.
22. Rogers, L. S., Knauss, Joanne, & Hammond, K. R. Predicting continuation in therapy by means of the Rorschach test. *J. consult. Psychol.*, 1951, **15**, 368–371.
23. Rosenberg, S. The relationship of certain personality factors to prognosis in psychotherapy. *J. clin. Psychol.*, 1954, **10**, 341–345.
24. Sandler, J. Rorschach content analysis: An experimental investigation. *Brit. J. med. Psychol.*, 1951, **24**, 180–201.
25. Silverman, H. The prediction of consciousness of conflict in the self from the Rorschach. Unpublished doctoral dissertation, Univer. of Michigan, 1952.
26. Steinzor, B. Rorschach responses of achieving and non-achieving college students of high ability. *Amer. J. Orthopsychiat.*, 1944, **14**, 494–504.

Journal of Consulting Psychology
Vol. 22, No. 5, 1958

The Relationship of Stimulus Ambiguity on the TAT to the Productivity of Themes[1]

Bernard I. Murstein
University of Portland

One of the problems confronting the clinician is the relationship of the ambiguity of TAT cards to the stories elicited. Medium-ambiguous pictures have been shown by Kenny (3) to educe the most personality-revealing stories. The present study sought to measure the relationship between ambiguity and the productivity of themes. A curvilinear relationship was hypothesized.

The rank order of 20 "male series" TAT cards obtained by Bijou and Kenny (1) for 28 college males, 23 college females, and total group of 51, was used to measure the ambiguity (number of possible interpretations) variable. The productivity variable was measured by the rank order of number of themes obtained by Eron (2) with the same cards for 150 male veterans of whom 50 were collegians. The ranks of both variables were converted to normalized scores on the assumption that both variables were normally distributed. Pearson r's and η's were obtained as well as chi squares and probability values for the differences between the two correlations.

Results of the correlation between the Eron group and the various combinations of the Bijou group were: Regression of productivity on ambiguity, Bijou males r .07, η .61, ($p < .10$); Bijou females r .03, η .43, ($p < .60$); Bijou total r .09, η .86, ($p < .001$). Regression of ambiguity on productivity, Bijou males r .07, η .95, ($p < .001$); Bijou females r .03, η .43, ($p < .60$); Bijou total r .09, η .73, ($p < .01$).

The results indicate the significant curvilinear relationship and nonsignificant linear relationship between ambiguity and productivity. The curvilinear relationship was such that a medium degree of ambiguity elicited maximum productivity, while both high and low ambiguity resulted in less productivity. The smaller degree of relationship found when using the Bijou female group was probably a function of the "male series" of cards and the difference in sex from the all-male Eron group.

A perusal of the TAT cards yielded the opinion that the moderately ambiguous cards were, for the most part, quite clear in so far as perception of the stimulus properties was concerned. The ambiguity appeared to reside in the question as to what emotions the characters were experiencing. Future research might well be directed towards investigating whether greater "personality revealingness" is obtained from vague stimulus figures or from clear stimulus figures whose feelings or activities are not readily apparent.

Brief Report.
Received May 26, 1958.

References

1. Bijou, S. W., & Kenny, D. T. The ambiguity values of TAT cards. *J. consult. Psychol.*, 1951, **15**, 203–209.
2. Eron, L. D. A normative study of the Thematic Apperception Test. *Psychol. Monogr.*, 1950, **64**, No. 9 (Whole No. 315).
3. Kenny, D. T. Transcendence indices, extent of personality factors in fantasy responses, and the ambiguity of TAT cards. *J. consult. Psychol.*, 1954, **18**, 345–348.

---

[1] An extended report of this study may be obtained without charge from Bernard I. Murstein, Psychology Department, University of Portland, Portland 3, Oregon, or for a fee from the American Documentation Institute. Order Document No. 5695, remitting $1.25 for microfilm or $1.25 for photocopies.

Journal of Consulting Psychology
Vol. 22, No. 5, 1958

Countertransference Effects in Psychotherapy[1]

Richard L. Cutler
University of Michigan

The therapist's personality has long been recognized as one of the most important variables in psychotherapy. Psychoanalytic writers have stressed repeatedly the importance of a personal analysis for the therapist, pointing out the deleterious effects upon the treatment which his unresolved conflicts may have. In spite of this emphasis in the theoretical literature and in practice, researchers have had little success in their attempts to investigate the influence of the therapist's personality upon his conduct of psychotherapy.

Two reasons for this lack of success are immediately apparent. The first, the extreme difficulty in the operational definition of psychoanalytic concepts, is common to all research on the psychoanalytic theory of personality. The second is unique to research in psychotherapy, particularly to that dealing with countertransference. It is the fact that most therapists, having training in at least the rudiments of personality theory and psychodiagnosis, do not readily drop their defenses to allow us to examine their conflicts and inner feelings, at least not short of the analytic couch.

Problem

The present study approaches the problem of the manifestation and effects of countertransference through a reinterpretation of the concept in terms of Bruner's theory of perception. In this way, some of the operational problems are overcome. In addition, by means of a rating device developed specifically for the purpose, it allows the identification of conflict areas in the therapist's personality, and the prediction of their effects upon the moment-to-moment interaction between patient and therapist.

Countertransference, as examined in this research, is defined as the transference reactions of the therapist to his patient. Fromm-Reichmann (8) defines transference in its most general sense as "the process of transferring to and repeating early patterns of interpersonal relatedness with present day partners." In its special application to the therapeutic relationship, she points out that it means transferring onto the therapist, with no or minimal basis in the realities of the situation, early experiences in interpersonal relatedness.

Fenichel (4) discusses transference in terms of impulses arising and being experienced which are inappropriate to the situation in which they arise. He emphasizes the tendency of the person experiencing transference reactions to attempt to reconcile the experience of the inappropriate impulses with the realities of the situation, and thereby to distort these realities to fit the demands of his own needs. He agrees with Freud (7) in terming the appearance of such distortions in the therapist as countertransference.

Herman Nunberg, speaking as a member of a panel on "Problems of Transference and Countertransference," dealt with the relationship between the unconscious needs of a person and his perceptions of reality. He says: ". . . the patient desires to find satisfactions for his needs in the person of the analyst, and attempts to change him into an object which will provide this satisfaction. He is frustrated because the analyst is not this object" (9, p. 26).

From this point of departure, Nunberg sketched the following picture: We carry within us images of the need-satisfying objects of old and try to find them in the outside world. We seek something in reality that is exactly like them and desire to bring about an identity between the perception of present objects and the images within us. The patient may unconsciously try to change reality so as to conform with the inner picture, or may seize upon some detail in the present reality which is similar to the desired object, and, aided by such detail, project his inner images upon the current reality. This process Nunberg sees as transference.

---

[1] This study was made possible by the support of U.S.P.H.S. Project M-516, "Analyses of Therapeutic Interaction," E. S. Bordin, Principal Investigator. The author wishes to thank all staff members whose cooperation made it possible.

We find a parallel formulation by Bruner (2) in the field of perception. Postulating a three-step process in perception, which consists of (*a*) hypothesis (set, Aufgabe, determining tendency); (*b*) input of information from the environment; and (*c*) confirmation or infirmation of the hypothesis, Bruner considers in detail the influence of need upon perception. Hypothesis strength, a central concept in Bruner's theory, is in part dependent upon the consequences which the particular hypothesis has in aiding the organism in the fulfillment of needs. The more basic the confirmation is to the carrying out of need-satisfying activity, the greater will be its strength. Bruner states three theorems which are contingent upon this concept of hypothesis strength:

1. The stronger an hypothesis, the greater its likelihood of arousal in a given situation.
2. The greater the strength of the hypothesis, the less the amount of appropriate information necessary to confirm it.
3. The greater the strength of an hypothesis, the more the amount of inappropriate or contradictory information necessary to infirm it (2, p. 126).

Thus it can be seen that strong need-satisfying hypotheses will tend to be confirmed on the basis of minimal appropriate information from the environment, a situation which is directly analogous to what Nunberg terms "seizing upon some detail in the environment in order to project inner wishes upon the current reality." We may conclude that the transference and countertransference phenomena are therefore special cases of perception being influenced by need.

The above formulation permits us to approach the problem of countertransference from the frame of reference of the influence of the therapist's needs upon his perception of his own and the patient's behavior in psychotherapy. If the formulation is valid, we should find certain systematic relationships between the needs and conflicts of the therapist and his ability to report objectively the behavior of himself and the patient in psychotherapy. When no conflict is present, and no defensive countertransference reactions are necessary, he should perceive and report this behavior with relative objectivity. On the other hand, when the stimulus situation is such that material impinges upon the therapist's needs and conflicts, we should find systematic tendencies to omit, distort, or overemphasize certain aspects of the behavior. This discussion leads to the first major hypothesis of the study: In reporting his behavior and that of the patient, the therapist will over- or underemphasize material related to his own needs and conflicts as compared to material which is related to relatively conflict-free areas in his personality.

If such effects do indeed result from the influence of the therapist's needs, it follows that there will be certain consequences for the process of therapy itself. One of these may be a rather gross misjudgment of what the patient's characteristic interpersonal relationships are like. In practice, however, steps are taken to prevent such gross errors by the therapist. By means of supervisory sessions and the utilization of projective tests, it is possible to check the therapist's judgments against less involved and more objective criteria. However, in the moment-to-moment interaction between therapist and patient, there is little immediate opportunity to control the distortions of the therapist. It is here that we may expect to find the clearest effects of the influence of the therapist's needs and consequent distortions upon the process of therapy.

Stekel (10) says, ". . . scotomata in the therapist due to unresolved infantile conflicts prevent him from dealing effectively with similar material when it is presented by his patient." This formulation is based upon extensive clinical experience and a profound knowledge of psychoanalytic theory, but it may be derived independently from the "need-influences-perception" frame of reference. Suppose that the patient exhibits behavior which touches directly or symbolically upon conflict areas in the therapist. This material will be threatening to the therapist, and immediate defensive operations will need to be undertaken to control the threat. This defensive operation may take the form of a perceptual distortion by the therapist of what the patient is doing. In addition, a portion of his attention and energies must necessarily be diverted away from the task of psychotherapy to the equilibrium-restoring process of defense. This will then become evident in a less adequate handling by the therapist of such conflict-relevant material. The above discussion leads to the second major hypothesis of the study: When behavior exhibited by the patient is similar to behavior which has been identified as conflictual for the therapist, the therapist's responses to this behavior will be judged to be significantly less adequate for therapeutic purposes than his responses to material which is relatively nonconflictual for him.

Procedure

Therapist conflict areas were identified by a method which consisted essentially of comparing therapist's self-ratings on certain personality traits with ratings made by judges who were attempting to describe the same therapist with regard to the same personality traits. Similar self-vs.-judge ratings have been used previously by other investigators, and the reliability of such ratings was sufficiently high to encourage their use in the present study (1, 5). The personality traits on which the therapists and judges were asked to make ratings were derived from a device called "The Circle" (6) which had been developed specifically for the purpose of categorizing interpersonal relationship materials of the kind found in therapy recordings, TAT stories, etc. The use of conflict and nonconflict areas based directly upon this coding system permitted a direct relating of the therapist need areas to the material collected in his therapy interviews and reports. An attempt was made to select words which would be descriptive of meaningful personality traits, and which would also adhere closely in meaning to the various points on the Circle. The sixteen adjectives finally selected are listed below: (Letters denote code category on the Circle.)

A. Dominating
B. Boastful
C. Rejecting
D. Punitive
E. Critical
F. Complaining
G. Suspicious
H. Apologetic
I. Submissive
J. Respectful
K. Dependent
L. Agreeable
M. Affiliative
N. Supportive
O. Generous
P. Advising

Each subject therapist was asked to place on a 19-point scale, from "most characteristic" to "least characteristic," each of the 16 traits as he felt it applied to himself in his day-to-day dealings with people. Nine or more judges, each of whom was personally well acquainted with the therapist he rated, carried out the same task, i.e., he rated each of the 16 traits as he felt it applied to the subject on the basis of his day-to-day personal and professional contact with the subject.

Conflict areas were identified as those traits which showed a significant disagreement [2] between the self-rating of the therapist and the judges' ratings. The assumption upon which this procedure rests is as follows: judges' ratings of a therapist are based upon how the therapist actually behaves in relation to each judge. If the therapist overrates or underrates himself on a particular trait, it is because he is unable to see himself as he really is. The reason for this lack of objectivity is conflict which the therapist has over the expression or nonexpression of the particular trait.

Using this technique, two types of conflict could be distinguished. The first, consisting of those traits on which the therapist overrated himself as compared to the judges' ratings of him, will hereafter be denoted as "plus-conflict." The second type, in which the therapist underrated himself, will be called "minus-conflict." Traits were judged to be relatively conflict-free for a subject when the judges' ratings clustered symmetrically about the therapist's self-rating. It is obvious that the total of the traits identified as conflict areas plus those identified as conflict-free or neutral will not account for all of the 16 behavior traits in each subject. In those cases where neither an identification of "conflict" nor "neutral" could be made, it was felt that the information provided by the therapist evaluation device was sufficiently equivocal not to permit any predictions to be made about these traits.

---

[2] The statistic used to establish significance of disagreement between subjects' self-ratings and judges' ratings was the "sign test." This is a nonparametric statistic which does not require the assumption of underlying normality of the population of judges' ratings on a particular trait. It merely requires that each judge's rating can be categorized as "higher" or "lower" than the subject's rating on any particular trait. Significance tables are available for this statistic and reveal that with nine judges, an 8–1 split is significant at the .05 level, and a 9–0 split is significant at the .01 level (3). In cases where the subject's rating on a trait tends toward the extreme of the rating scale, the sign test becomes increasingly less applicable, so that trait judgment discrepancies which were the result of the subject rating himself at one of the extremes of the scale were not used as conflict areas.

In all, ten subjects were pretested in order to identify conflict areas. There was no systematic tendency for any single trait or group of traits to be identified as "conflictual" across the subjects, nor were any of the traits systematically overrated or underrated across all subjects.

Two therapists were selected for the major portion of the study. The major factor in the selection was the requirement that they have low similarity insofar as identified neutral and conflict areas, although considerations of time, willingness to cooperate, and availability of patients were also involved.

Therapist 1, who had had three years of graduate training in clinical psychology, more than 300 hours' experience as a therapist, and had completed a personal psychoanalysis, tape recorded three consecutive therapy interviews with each of three of his patients in a college student counseling center. Therapist 2 was in his second year of graduate training in clinical psychology, had had less than 50 hours' experience as a therapist, and had had no personal psychotherapy. He recorded four consecutive sessions with each of two patients, who were being seen in a Veteran's Administration facility.

In addition to tape recording the interviews, each therapist dictated a detailed account of what he thought had transpired during the session immediately following each therapy hour. These reports were not less than 500 words long, and included his impressions of all of his own and the patient's behavior during the hour. The tape-recorded interviews and dictated reports were transcribed verbatim according to a standard set of transcription rules, which permitted identification of pauses, changes in rate of speech and vocal expressions of emotion, as well as the actual verbal content.

The interviews were coded on the Circle, and by summing across all the interviews of a given therapist, the total number of actual manifestations by the therapist and by the patient of each of the 16 interpersonal behavior categories of the Circle was ascertained. This summation was assumed to be the *objective* picture of the behavior of the therapist and his patients during the series of interviews.

A similar coding was carried out on the therapist reports. By summing the number of occurrences of each of the 16 behaviors across all the reports of a given therapist, a picture was obtained of each therapist's *perception* of his own and his patients' behavior.

The coding reliability of the categorization scheme for interview and report material was established by spot-checking interviews which were randomly selected from the total group. One judge coded the entire set of interviews, another coded the entire set of reports, and a third coded three randomly selected interviews and three randomly selected reports in order to establish interjudge reliability. Since the usual reliability statistics are not applicable to codings of this kind, reliability was estimated simply by counting the number of exact agreements in coding a given statement or response, as well as partial agreements and disagreements.[3] For the three interviews, exact agreement ranged from 62% to 71%, partial agreement from 24% to 33%, and disagreement from 5% to 10%. The reliability estimates were somewhat lower for the report material, but this was partially due to the fact that the reports were not coded sentence by sentence, and there was no requirement that each sentence be coded. It was felt that under these circumstances, it was not meaningful to express agreement in terms of percentages, and product moment correlations were run between the total numbers of ratings of each trait by the two judges. The results of three such tests yielded *r*'s of .75, .81, and .84, and were interpreted as indicating that the reports could be reliably coded by independent judges. In view of the fact that only the total number of appearances of each trait in the reports entered into the testing of the hypotheses, it was felt that demonstrating point-to-point agreement among the judges was not necessary so long as the total number of codings for each trait agreed to the extent indicated above.

---

[3] Since a single response by therapist or patient might contain one, two, or more Circle category codes, partial agreement was possible. The rule of thumb established for determining complete agreement, partial agreement, and disagreement was as follows: only when all codes for a given response corresponded exactly was complete agreement indicated; when half or more corresponded, it was considered a partial agreement; and when less than half corresponded, it was considered a disagreement.

Specific predictions were made regarding the relative accuracy of the therapist's reports of behavior which had been identified as conflictual as compared to the reports of behavior identified as neutral. For the therapist's reports of his own behavior, it was predicted that he would systematically overreport those occurrences of behavior in himself which had been identified as plus-conflict, and underreport those which had been identified as minus-conflict.

Concerning the therapist's perception of the behavior of his patients, it was predicted that his reports, compared with the actual number of occurrences of a given type of behavior, would differ significantly from his reports regarding neutral material.

In order to test the hypothesis concerning the adequacy of the therapists' handling of conflict vs. neutral material, each therapist response was coded by an independent judge as being either "task-oriented" or "ego-oriented." By definition, task-oriented behavior was any behavior on the part of the therapist which tends to produce or facilitate in the patient a flow of therapeutically relevant conversation, while ego-orientation involves therapist behavior which departs from this task for the apparent purpose of expressing the therapist's own needs. Thus, a consistent tendency on the part of the therapist to use ego-oriented mechanisms will result in a lessened therapeutic efficiency and a consequent slowing down of the patient's progress.

We may consider ego-oriented behavior as a form of defensive activity on the part of the therapist. So long as the patient's productions do not impinge upon conflict areas, there is little need for the therapist to indulge in such defensive activity, and he will adhere to the task of psychotherapy. However, when the

Table 1

Results: Therapist 1

			Therapist self-report vs. actual occurrence				Therapist report of pt. vs. actual occurrence			
Code	Trait	Conflict	Actual	Reported	Prediction	Significance	Actual	Reported	Prediction	Significance
A	Dominating	Neutral	25	17	—	—	7	5	—	—
B	Boastful	—	1	1	None	—	64	60	None	.01[a]
C	Rejecting	—	2	8	None	.001[a]	8	49	None	.001[a]
D	Punitive	Plus	5	11	Overreport	.01	4	18	Disparity	.001
E	Critical	Plus	1	12	Overreport	.001	1	39	Disparity	.001
F	Complaining	Plus	1	0	Overreport	—[b]	96	47	Disparity	.50[c]
G	Suspicious	Plus	4	13	Overreport	.001	5	28	Disparity	.001
H	Apologetic	—	5	2	None	—	128	74	None	—
I	Submissive	Plus	2	4	Overreport	.15[c]	11	16	Disparity	.01
J	Respectful	—	185	19	None	.001[a]	35	33	None	.01[a]
K	Dependent	Plus	0	4	Overreport	.02	26	31	Disparity	.01
L	Agreeable	Minus	77	8	Underreport	.001	115	53	Disparity	.50[c]
M	Affiliative	Minus	11	1	Underreport	.05	6	35	Disparity	.001
N	Supportive	—	14	39	None	.001[a]	0	0	None	—
O	Generous	Minus	94	70	Underreport	.10	6	3	Disparity	.50[c]
P	Advising	Neutral	191	99	—	—	226	116	—	—

[a] Significant disparity where no prediction was possible.
[b] *N* insufficient to permit significance test.
[c] Unverified prediction.

Table 2

Results: Therapist 2

			Therapist self-report vs. actual occurrence				Therapist report of pt. vs. actual occurrence			
Code	Trait	Conflict	Actual	Reported	Prediction	Significance	Actual	Reported	Prediction	Significance
A	Dominating	Minus	23	3	Underreport	.01	0	6	Disparity	.01
B	Boastful	Plus	1	0	Overreport	—[b]	62	0	Disparity	.001
C	Rejecting	Plus	4	15	Overreport	.01	4	21	Disparity	.001
D	Punitive	Minus	13	0	Underreport	.001	7	9	Disparity	.10
E	Critical	Plus	0	0	Overreport	—[b]	0	15	Disparity	.001
F	Complaining	—	4	2	None	.50	131	48	None	.01[a]
G	Suspicious	Plus	1	8	Overreport	.01	4	28	Disparity	.001
H	Apologetic	Plus	14	6	Overreport	.40[c]	46	54	Disparity	.05
I	Submissive	—	3	5	None	.50	7	13	None	.01[a]
J	Respectful	Minus	119	4	Underreport	.001	87	25	Disparity	.001
K	Dependent	Plus	0	1	Overreport	—[b]	5	15	Disparity	.01
L	Agreeable	Plus	29	54	Overreport	.001	97	34	Disparity	.001
M	Affiliative	—	14	16	None	.05[a]	9	24	None	.01[a]
N	Supportive	Neutral	128	77	—	—	3	3	—	—
O	Generous	Minus	42	14	Underreport	.01	2	3	Disparity	.50[c]
P	Advising	Neutral	248	164	—	—	218	163	—	—

[a] Significant disparity where no prediction was possible.
[b] N insufficient to permit test.
[c] Unverified prediction.

patient presents behavior which does touch the therapist's conflicts, the therapist might be expected to resort to ego-oriented behavior for the purpose of defending himself against the anxiety which the patient's productions arouse. This discussion leads to the following prediction: there will be a significant association between the appearance of ego-oriented behavior in the therapist, and the appearance in the patient's previous statement of material which has conflict-relevance for the therapist.

The coding reliability of the task-ego categorization of therapist responses was established by having an independent judge code three randomly selected interviews response-by-response for either task or ego-orientation. This independent judge agreed in 78%, 83%, and 85% of the responses with the judge who coded the entire set of interviews. These figures far exceed chance expectation, and are considered to be sufficient evidence for the reliability of the task-ego categorization scheme.

Results

The results of the study bearing on the prediction of the accuracy of the therapists' reporting of what transpired in the therapy sessions are presented in Tables 1 and 2.

An inspection of Tables 1 and 2 indicates the general substantiation of the hypotheses concerned with the accuracy of the therapist's reporting where conflict and neutral material was compared. Of a total of 40 predictions made, 28 were completely supported, 2 others showed trends in their favor, 4 could not be tested because of an insufficient amount of material, and 6 were completely without verification. In addition, a total of 10 significant disparities were found for which we had been unable to make predictions.[4] The failure of the therapist evaluation device to predict in every case where distortions would occur is admittedly a weakness of the instrument. However, the improvement in prediction which it does permit is extremely encouraging.

The results of the second major portion of the study are presented in Tables 3 and 4.

---

[4] In every case, significance was tested by means of the chi-square test of association or Fisher's exact test. The number of occurrences and reports of those traits which had been identified as neutral was used as a standard against which to compare the frequency of occurrence and reported occurrence of each of the other traits.

The specific prediction that therapist responses following patient responses containing interpersonal interactions, which had been previously judged to be conflictual for the therapist, would tend to be ego- rather than task-oriented, finds strong support in the data.

Discussion

At the very least, we may conclude that the general formulations concerning the manifestations and effects of countertransference are substantiated. In addition, it is clear that the therapist's reports of his own and his patients' behavior in therapy are not to be trusted as accurate. In 25 out of 27 cases of overreporting, there was distortion not only on the basis of the comparisons with the neutral traits, but also on an absolute basis. That is, the therapists reported more of the behavior than actually occurred. Varying levels of experience in therapy and of increased self-awareness brought about by personal analysis do not seem to make a difference. We may also conclude, with somewhat less confidence, that the therapist evaluation device provides us with a means of predicting those areas in which distortions will occur, and in the case of the therapist's self-reports, the direction that these distortions will take.

It is unfortunate that these personality variables are not more readily translatable into terms which can be applied to psychoanalytic theory. It is possible that there is a relationship between certain tendencies to over- or underemphasize a behavior trait in one's self or in others, and the classical defense mechanisms described in psychoanalytic literature. The tendency of Therapist 1 to project his own problem areas into his patients' behavior seems quite clear. Tendencies to underestimate the frequency of appearance of certain traits in one's self may parallel the mechanisms of denial and repression, and to overestimate them, the mechanisms of intellectualization or isolation. It must be admitted that the data of the present study are insufficient to allow a definitive statement about this issue. The problem of relating overt behavior types to the theory of defense mechanisms is one which requires further investigation. The methodology of this study seems to offer a promising means for such an investigation.

Table 3

Therapist 1

Category	Conflict	Nonconflict	Total
Task	159	202	361
Ego	71	52	123
	230	254	484

Note.—Chi square 6.29; $p < .02$; df 1.

Table 4

Therapist 2

Category	Conflict	Nonconflict	Total
Task	58	115	173
Ego	185	76	261
	243	191	434

Note.—Chi square 59.5; $p < .001$; df 1.

It seems clear that the appearance in the patient of behavior which is conflict relevant for the therapist prevents the therapist from functioning at maximum efficiency. In addition, there seems to be a definite relationship between the amount of experience and/or self-insight which the therapist has, and his tendency to show task-oriented, rather than ego-oriented behavior. It appears that even though both therapists' perception of their patients is disturbed by their own conflicts, Therapist 1 is able to make use of his experience to behave in a more appropriate and effective manner in the actual process of psychotherapy. Apparently, training and supervision do pay dividends in terms of increasing the benefits of psychotherapy to the patient, if the proportion of task vs. ego-oriented statements by the therapist is accepted as a rough criterion of therapeutic effectiveness.

In addition to the relevance which this study has for the selection and assignment of patients to therapists whose personalities would be least limiting in the relationship with a given patient, and the importance which it has in regard to the confirmation of analytic formulations concerning the manifestations, sources, and effects of countertransference, the possibility of further investigation of psychoanalytic theory through a reformulation in terms of concepts from general psychology should not be overlooked. Such attempts to reconcile evidence from two fairly separate spheres of psychology may permit us to move more rapidly toward a truly general theory of behavior.

Summary

This study was concerned with the effects of countertransference reactions in the therapist upon his perception of his own and his patients' behavior in psychotherapy, and upon his effectiveness in dealing with patient material which impinged upon his own areas of conflict. Based upon a reformulation of the psychoanalytic concept of transference in terms of Bruner's theory of perception, two hypotheses were developed for test. The first postulated a systematic relationship between the therapist's conflicts, and his tendency to over- or underreport the occurrence of similar behavior in himself and his patient in psychotherapy. The second hypothesized that the therapist's therapeutic handling of material which was conflict-relevant for him would be less adequate than his handling of material which was relatively conflict-free.

Therapist conflict areas were identified by means of a specially developed rating scale based upon adjectives (traits) derived from the "Circle" interpersonal mechanism coding scheme developed by Freedman et al. The appearance of significant disparities between the therapist's rating of himself and judges' ratings of him was assumed to indicate the presence of conflict. Two therapists, whose identified conflict areas were dissimilar, tape recorded a series of interviews and, immediately following each therapy session, dictated a detailed account of what they perceived to have transpired during the interview. These interviews and reports were then coded by means of the Circle. The interview codings were assumed to represent an objective account of what had transpired, and were used as a standard against which to compare the therapist's reports. In addition, each therapist response was coded as being task-oriented or ego-oriented, and the latter was assumed to represent a less adequate handling of the patient than the former.

Of a total of 40 predictions made regarding the therapist's tendency to distort his reports of need-relevant behavior, 28 were clearly confirmed, 2 found partial support, 4 could not be tested because of an insufficient number of occurrences of the behavior in question, and 6 were not supported. In addition, there was a significant association between the appearance, in the patient's statement, of material which impinged upon the conflict areas of the patient and the judged inadequacy of the therapist's immediately following response.

The results were interpreted to offer strong support, both for the validity of the Freudian formulations concerning the effect of countertransference and for the adequacy of the translation of these formulations into Bruner's terms.

The feasibility of the objective study of the therapeutic relationship was pointed out, and the potential fruitfulness, for future research, of additional translations of Freudian concepts into the terms of general psychology was indicated.

Received August 29, 1957.

References

1. Aronson, M. A study of the relationships between certain counselor and client characteristics in client-centered therapy. Unpublished doctoral dissertation, Pennsylvania State Univer., 1952.
2. Bruner, J. Personality dynamics and the process of perceiving. In R. N. Blake & G. V. Ramsey (Eds.), *Perception: An approach to personality*. New York: Ronald, 1951. Pp. 121–147.
3. Dixon, W. J., & Massey, F. J. *Introduction to statistical analysis*. New York: McGraw-Hill, 1951.
4. Fenichel, O. *Problems of psychoanalytic technique*. Albany: Psychoanalytic Quarterly, 1942.
5. Filer, R. N. The clinician's personality and his case reports. Unpublished doctoral dissertation, Univer. of Michigan, 1951.
6. Freedman, M., Leary, T., Ossorio, A., & Coffey, H. The interpersonal dimensions of personality. *J. Pers.*, 1950, **20**, 143–162.
7. Freud, S. *Collected papers, Vol. II*. London: Hogarth, 1950.
8. Fromm-Reichmann, Frieda. *Principles of intensive psychotherapy*. Chicago: Univer. Chicago Press, 1951.
9. Nunberg, H. Problems of transference and countertransference: A panel discussion. *Bull. Amer. Psychoanal. Ass.*, 1950, **6**, 24–27.
10. Stekel, W. *Technik der analytischen Psychotherapie*. Bern: Huber, 1938.

Journal of Consulting Psychology
Vol. 22, No. 5, 1958

Social Desirability and Self-Ratings of Intakes, Patients in Treatment, and Controls[1]

H. J. Wahler

Veterans Administration Hospital, Knoxville, Iowa

A number of recent studies (1, 2, 3, 7, 11) have consistently demonstrated a strong relation between self-descriptions and rated social desirability of personality items. Comparable correlations have been obtained with different item content, scaled and nonscaled social desirability weights, and various techniques such as true-false questionnaires, self-ratings, and *Q* sorts. The stability of certain social desirability stereotypes and their relative independence of such variables as age, education, sex, socioeconomic status, and cultural differences have been shown by Klett (8, 9) and others (1).

All of the findings cited, involving self descriptions, were based on the responses of student samples. These studies show that a high degree of correspondence between averaged self descriptions and item desirability is a general response characteristic of such groups. Rosen (11) obtained results identical to those of Edwards (1) with his student *groups*. However, he found that when the self appraisals of *individuals* were correlated with social desirability, the obtained coefficients ranged from negative to positive values, with a preponderance on the positive side. These results indicate that not all people give self descriptions which agree positively with item desirability.

To account for the relationship which he found, Edwards suggested two hypotheses. One was that people either consciously or unconsciously tend to describe their personality traits in such a way as to "look good." They may accomplish this by indicating that socially desirable items are characteristic of them and that undesirable items are not, regardless of the validity of such assertions. The second assumption was that personality traits judged desirable are in actuality more typical of members of a culture than are undesirable traits, and this is reflected in the averaged self-descriptions of nondeviant groups.

Considering only the correspondence of subjects' (*S*s') traits (which could be defined operationally in terms of specified personality assessment procedures) and self-descriptions with social desirability, Edwards' hypotheses imply two general types of respondents: (*a*) *S*s whose traits (actually a selected sample) are not in agreement with social values but whose self-appraisals correspond with item desirability and (*b*) *S*s whose traits correspond with social desirability and whose self-descriptions reflect this. Rosen's findings provide a basis for assuming two additional types of respondents: (*c*) *S*s whose personality traits do not correspond with social values and whose self-descriptions reflect this, and the logical possibility of (*d*) *S*s whose personality traits correspond with social values but whose self-descriptions are not positively related to item desirability[2] (malingerers?).

---

[1] The author wishes to extend his appreciation to S. J. Williamson and Harold Bechtoldt for their kind assistance in obtaining student ratings and to Lynn Roberts, Margaret Walsh and Marvin Graber for their generous help in rating patients and scheduling patients for testing. The author is also very grateful to Rowena Rash for her painstaking assistance with the major portion of the computations necessary for this study.

[2] While these four suggested classifications are expressed in categorical terms to facilitate exposition, it is more realistic to assume that all of the factors are a matter of degree, i.e., personality traits of individuals are more or less in agreement with social values as are the self-appraisals of groups or individuals.

Since self-descriptive techniques are widely used in clinical practice, it is important to investigate the degree of correspondence between the self-appraisals of clinical *S*s and item desirability. Edwards and others have shown that if the social desirability weights of a set of personality items are known, approximately 75% of the variance in self-appraisals of nonclinical groups can be predicted from such weights. If it were found that the self-appraisals of clinical groups corresponded as highly with item desirability and could not discriminate clinical from nonclinical *S*s on any other basis, the clinical usefulness of such techniques would be subject to even more serious conjecture than is currently the case (4, 5, 6).

With clinical *S*s, particularly those who are receiving psychotherapy, operationally defined procedures may be readily used to obtain appraisals of certain of their personality traits by therapists. The correspondence between therapists' appraisals of patients and social desirability and the correspondence between clinical *S*s' self-descriptions and item desirability may provide an objective basis for classifying clinical *S*s in terms of respondent characteristics such as those suggested above.

The clinical literature contains numerous descriptive accounts of various traits judged to be characteristic of patients. Many of these traits, when sufficiently pronounced, are considered socially undesirable, i.e., hostility, dependency, immaturity, suspiciousness, self-preoccupation, poor interpersonal relationships, etc. If such traits were judged to be typical of clinical *S*s, it might be anticipated that therapists' appraisals would not correlate significantly with social desirability. If, however, clinical *S*s tend to deny undesirable traits which their therapists attribute to them and give self-descriptions corresponding primarily to social desirability stereotypes, this could readily be interpreted as defensiveness and/or lack of insight. Under these circumstances, clinical groups could be classified as *S*s whose traits are not in agreement with social values but who give self-appraisals corresponding with item desirability.

An investigation of these factors in conjunction with clinical *S*s cannot ignore the possibility that psychotherapy may affect the self-descriptive characteristics of such people. It is widely held by many clinicians that, among other effects, psychotherapy tends to reduce defensiveness and/or increase insight or realistic self-awareness. If therapy has this effect, the self-descriptions of people receiving therapy (patients) should be less strongly related to item desirability (and more strongly related to therapists' estimates of their traits) than would be the case with people seeking such assistance for the first time (intakes). This, of course, would not hold if the self-descriptions of intakes were in no way indicative of defensiveness. Considering the repeated emphasis in the literature on the observed defensiveness of people undertaking therapy, particularly at the outset, it seems probable that their self-descriptions may also reflect this tendency.

While this investigation is primarily concerned with the relationships between social desirability and self-descriptions of clinical *S*s and therapists' appraisals, there are other characteristics of the self-descriptions of clinical *S*s that may be studied with the obtained material. One is the level at which they ascribe various traits to themselves relative to nonclinical groups. Such an analysis is important since it is possible to obtain the same degree of correlation between self-descriptions and item desirability for two different groups and still find that one group scored at a significantly higher level than the other. Another property concerns the correspondence between self-descriptions obtained from various groups and what may be termed normative estimates of the traits of a criterion group. In this case, a check on the "validity" of the self-descriptions of clinical groups can be made by determining the extent to which they correspond with therapists' appraisals of patients. If intakes and patients are describing themselves with any validity, their self-appraisals would be expected to correlate significantly with therapists' appraisals. On the other hand, the self-descriptions of nonclinical *S*s would not be expected to correspond as highly with the judged traits of clinical *S*s.

One purpose of this study is to compare the degree of correspondence between self-ratings and independently obtained social desirability weights for a set of personality items found with intakes, patients, and student and nonstudent control groups. A second purpose is to compare the coefficients obtained by correlating averaged therapists' ratings of a group of patients with the self-ratings obtained from the above groups. The above groups will also be compared with regard to the relative levels at which they ascribe clinically meaningful traits to themselves which are classified as undesirable, slightly undesirable, and desirable.

Klett's (9) findings with groups of *hospitalized* psychotics suggest that there may be some systematic differences in judged social desirability between clinical and nonclinical groups. Since differences in conceptions of social values could affect self-descriptions, the agreement among social desirability ratings obtained from students and intake and patient samples from an outpatient clinic will be determined.

Procedure

Subjects

Two different groups of male clinical *S*s were used. The *patient* group consisted of 42 veterans receiving psychotherapy at an outpatient mental hygiene clinic. Only patients who were not judged to be grossly disturbed or psychotic were asked to participate. These people had received an average of 84.83 one-hour individual interviews; the number of interviews ranged from 12 to 392. Patients gave social desirability and self-ratings in counterbalanced order. The *intake* group was composed of 25 veterans in the process of applying for treatment at the same clinic. All of these *S*s gave self-ratings; 6 were randomly selected to give social desirability ratings.

Four different student groups gave *social desirability ratings*. They were comprised of 62 male and 29 female students in introductory psychology classes at the State University of Iowa (SUI) and 6 male and 27 female students attending a night course in abnormal psychology at Drake University. *Self-ratings* were obtained from a group of 27 SUI male students and 30 male veterans receiving outpatient medical treatment for nonpsychiatric conditions at a Veterans Administration outpatient center. The veteran group was carefully screened to exclude *S*s with records of neurological or psychiatric diagnoses or complaints. None of the *S*s were older than 55 or younger than 18.

Self-rating Inventory

The inventory used to obtain self-ratings and social desirability ratings consisted of 44 items pertaining to characteristics commonly regarded as important clinical variables such as anxiety, hostility, sexual difficulties, dependency, suspiciousness, poor interpersonal relations, etc. All self-ratings were obtained with a 9-point scale. The extremes were anchored "not at all like me" and "beyond question very much like me." Social desirability ratings were also obtained with a 9-point scale. Extremes were anchored "definitely very highly socially acceptable" and "definitely very highly socially *un*acceptable." All *S*s were oriented to estimate how desirable the various traits would be in general to other people: Rosen's (11) "perceived social desirability."

Therapists' Ratings of Patients in Treatment

The 42 patients in treatment were being seen by four therapists during the time data was collected; 14, 11, 9, and 8 patients, respectively, were in treatment with different therapists. Therapists rated all patients who were serving as *S*s on the items of the self-rating inventory. The same nine-point scale was used except that the pronoun "him" was substituted for "me" in all anchoring phrases. Therapists also rated the items for social desirability.

Results

Social desirability ratings of the student groups were pooled since all intercorrelations were high ($\bar{r} = .922$) and there were no significant group differences.

The correlational matrix was computed for the averaged social desirability ratings of the combined student group ($N = 124$), intakes, patients, and therapists; the average r was .920. A test of over-all homogeneity was nonsignificant;[3] none of the group means differed significantly. These analyses indicate that the social desirability ratings obtained from stu-

---

[3] The results of any statistical test will be termed significant if the corresponding p value is .05 or less and nonsignificant if p is greater than .05.

Table 1

Correlations of Self-Ratings with Social Desirability Ratings for Controls, Intakes, Patients, and Therapists' Ratings of Patients

	Self-ratings				
Social desirability ratings	SUI Males ($N = 30$)	Outpatient veterans ($N = 27$)	Intakes ($N = 25$)	Patients ($N = 42$)	Therapists' ratings of patients ($N = 42$)
Students ($N = 124$)	.755*	.753	.661	.373	.008
Patients ($N = 42$)	.678	.696	.620	.337	−.016
Intakes ($N = 6$)	.708	.723	.526	.242	−.093
Therapists ($N = 4$)	.648	.693	.521	.211	−.103

* Coefficients based on averaged ratings of 44 items; $r = .385$ significant at .01 level; $r = .298$ significant at .05 level.

dents, intakes, patients, and therapists are highly comparable.

The averaged *self-ratings* obtained from outpatient veterans, SUI students, treatment and intake patients, and therapists' ratings of patients were correlated with mean social desirability ratings given by four different groups. These coefficients are presented in Table 1. None of the correlations in any given column differ significantly. Therefore, only r's in the first row of Table 1 are compared.

Correlations between judged social desirability and self-ratings obtained from the student and nonstudent groups are nearly identical. The self-ratings of these groups have about 57% common variance with item desirability. This value may be contrasted with 44% and 14% common variance with item social desirability in the self-ratings of intakes and patients, respectively. Therapists' ratings of patients were not significantly related to social desirability.

Table 2

Intercorrelations Among Mean Self-Ratings of Controls, Intakes, Patients, and Therapists' Ratings of Patients

	SUI Males	Intakes	Patients	Therapists' ratings of patients
Outpatient veterans	.817*	.710	.425	.085
SUI Males		.687	.514	.154
Intakes			.811	.422
Patients				.724

* Coefficients based on averaged ratings of 44 items; $r = .385$ significant at .01 level; $r = .298$ significant at .05 level.

While the correlation between self-descriptions and item desirability was slightly lower for intakes than for controls, the difference is short of significance. The correlation between self-descriptions and item desirability for patients is considerably less than for intakes and the two control groups.

Table 2 contains the matrix of intercorrelations among the averaged self-ratings of outpatient veterans, SUI students, patients, intakes, and therapists' ratings of patients. The test of homogeneity of r's was highly significant. Self-ratings of patients and intakes correlate significantly with therapists' ratings of patients; the self-ratings of both control groups are not significantly related to these values. Self-ratings of patients correlate significantly higher with therapists' ratings than do those of intake[4] or control groups. The self-ratings of intakes correlate significantly higher with therapists' ratings than do those of outpatient veterans but not SUI students.

---

[4] Therapists rated the 42 patients in treatment. It is assumed that idiosyncratic differences among individual patients and therapists were averaged out and that these pooled values are generally representative of traits manifested by patients seen at this clinic. However, it may be questioned whether intakes' self-ratings might not have corresponded more highly with therapists' ratings if they had been the individuals rated. To investigate this possibility, the averaged self-ratings obtained from 25 different intakes on 36 of the 44 items were correlated with ratings done by social workers who saw them for purposes of obtaining a case history. Social workers' ratings of intakes correlated .46 with their mean self-ratings. Although not an entirely satisfactory comparison since therapists had more opportunity for observation, this adds support to the finding that intakes' self-ratings do not correspond as highly with observers' ratings as do those of patients.

It may also be noted that self-ratings of the intake and patient groups correlate .811 with each other. Self-ratings of the two control groups correlate at about the same level (r = .817). The self-ratings of intakes or patients correspond to a lesser degree with the self-ratings of either control group; these r's range from .425 to .710.

The average levels at which the various groups rated themselves were compared. To make these comparisons, the 44 items were categorized as undesirable, slightly undesirable, and desirable.[5] The mean level of the ratings on items in these three categories was obtained for each *S* and for therapists' ratings of patients. Analysis of variance (10) based on these scores yielded highly significant effects for groups,[6] social desirability, and interaction. The mean ratings of the four groups on items in the three social desirability categories are presented in Table 3.

Tests of the significance of differences between mean self-ratings of the different groups within each category were made. The results show that patients rate items classified as undesirable and slightly undesirable at a significantly higher level than intakes. It was also found that both clinical groups tend to rate undesirable and slightly undesirable items at significantly higher levels than controls. Intakes and patients also rate desirable items at a higher level than controls; however, this difference is not significant. The best discrimination between controls and intakes or patients is obtained with ratings in the slightly undesirable category. Ratings in this category correctly identified 78% of intake and control *S*s and 85% of patients and controls.

Table 3

Mean Self-Ratings on Items Classified as Undesirable, Slightly Undesirable, and Desirable for Controls, Intakes, Patients, and Therapists' Ratings of Patients

Type of item	Therapists' ratings of patients	Patients	Intakes	Outpatient veterans
Undesirable (16 Items)	4.90	4.31	3.70	2.91
Slightly undesirable (20 Items)	6.05	5.77	5.14	3.88
Desirable (8 Items)	4.73	5.40	5.58	5.27

Discussion and Conclusions

The findings show that the self-ratings of intakes correlate with item desirability to about the same degree as the self-ratings of student and nonstudent controls. The self-descriptions of patients show only slight correspondence with cultural desirability stereotypes.

In the case of intakes, however, and to a much greater extent with patients, there is evidence that their self-descriptions are to some degree "valid." It was shown that intakes do attribute clinically meaningful characteristics to themselves to a greater extent than controls. Their self-descriptions correlated significantly with therapists' ratings while the ratings of nonclinical *S*s did not; their self-ratings correlated significantly more highly with those of patients in treatment than did ratings of controls; and intakes rated themselves at significantly higher levels than controls on items classified as undesirable and slightly undesirable. To this extent, their self-descriptions appear to contain "valid" variance. With patients, the evidence is much more strongly in this direction.

These findings, plus the fact that therapists' ratings of patients' traits did not correlate with social desirability, indicate that patients may be readily classified as *S*s whose personality traits do not correspond with social values and whose self-descriptions reflect this. It is apparent that the self-descriptions of patients cannot be as readily predicted from item desirability weights as appears to be the case with nonclinical *S*s. If certain traits of intakes, as a group, are comparable to those of patients, and hence do not correspond with social values, intakes would be classified as *S*s whose traits do not agree with social values but whose self-descriptions correspond with item desirability. Nevertheless, the self-descriptions of intakes can be reliably discriminated from those of nonclinical *S*s, since intakes rate undesirable and slightly undesirable traits at a significantly higher level than controls, and the self-ratings of intakes correlate significantly higher with therapists' ratings of patients. These findings lend support to the assumption that self-descriptive techniques have potential clinical and experimental utility.

---

[5] These classifications were based on the pooled social desirability ratings of students ($N = 124$). Arbitrary cutting points of mean values less than four, four to five, and greater than five were used to define the categories undesirable, slightly undesirable, and desirable. The number of items in each classification is 16, 20, and 8 respectively.

[6] The same analysis was done comparing outpatient veteran controls with SUI students. Since this analysis showed that there were no significant differences between these two groups, only scores from one group were included in this analysis. The outpatient veteran group was selected since it is more comparable to the veteran clinic groups in terms of age, education, socioeconomic level, and military experience than the students.

Comparing patients with intakes, the findings show that the self-descriptions of people receiving therapy correspond much less with item desirability than is the case with people seeking help who have not experienced psychotherapy. It was also shown that patients' descriptions correspond more closely with the appraisals of therapists than do those of intakes. These differences between patients and intakes are in agreement with the assumption that psychotherapy has the effect of reducing defensiveness and increasing insight or realistic self-awareness. The additional finding that the number of therapy interviews is significantly related[7] to the level at which patients ascribe undesirable characteristics to themselves is also in accord with this view.

Among both intake cases and patients in treatment, there were individuals who avoided ascribing undesirable or slightly undesirable traits to themselves. As mentioned earlier, a marked tendency to give self-descriptions which correspond primarily to cultural stereotypes of desirability, and to deny clinically meaningful traits judged undesirable among people sufficiently maladjusted to seek or participate in therapeutic assistance, suggests considerable defensiveness and/or lack of insight on their part. At least the probability is high that they are failing to provide personally relevant communication.

The extent to which such noncommunicative tendencies are related to various prognostic criteria has not been explicated. Questions may be raised, such as, do people seeking help with personal problems who are noncommunicative at the outset constitute poor therapeutic risks? Are they more likely to terminate therapy prematurely or fail to show signs of profiting from continued therapy than persons who are initially more communicative? The study of Rubinstein and Lorr (12) suggests that some factor such as noncommunicativeness may be related to premature termination. These authors found that ". . . [remainers] are more dissatisfied with themselves, feel worse, and see themselves as having poorer interpersonal and overall adjustment than terminators."

Since it was shown that patients receiving therapy were typically more inclined to ascribe undesirable and slightly undesirable traits to themselves than were intakes, the question may also be raised as to whether prognosis is more unfavorable for those patients who fail to show an increase in personally relevant communication after a certain number of interviews.

The evidence obtained in this study, while preliminary, suggests that scores based on the level at which intakes and patients rate themselves on clinically meaningful items classified in terms of social desirability values may provide useful indices of their ability or willingness to communicate personally relevant information. These measures may in turn have potential utility in evaluating response to therapy as well as possible selective and predictive properties.

Summary

The purpose of this study was to compare the relationships between judged social desirability and self-ratings obtained from patients in treatment at an outpatient clinic, intakes, and student and nonstudent control groups. Another purpose was to correlate the self-descriptions of these various groups with therapists' ratings of patients.

---

[7] While not high, the linear correlation was .39 ($p < .025$); eta was .66, but the deviation from linearity was not significant. If *S*s who had received no therapy had been included, these coefficients would have been larger.

The findings showed that the self-ratings of intakes correlated with social desirability weights to about the same degree as the self-ratings of controls. The correlation between the self-ratings and item desirability obtained with patients was significantly lower than the correlations obtained with intakes or controls.

Self-descriptions of both clinical groups correlated significantly with therapists' ratings of patients, while the self-ratings of the two control groups did not.

With items classified as undesirable, slightly undesirable, and desirable, it was shown that patients rated themselves significantly higher than intakes, and intakes rated themselves significantly higher than controls on undesirable and slightly undesirable items. Items in the slightly undesirable classification gave the best discrimination between clinical groups and controls.

The possibility of using scores based on items classified as undesirable and slightly undesirable as indices of personal communicativeness was discussed. It was suggested that such scores deserve further study as predictive and evaluative measures with regard to therapy.

Received September 3, 1957.

References

1. Edwards, A. L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.*, 1953, **37**, 90–93.
2. Edwards, A. L., & Horst, P. Social desirability as a variable in Q technique studies. *Educ. psychol. Measmt*, 1953, **13**, 620–625.
3. Edwards, A. L. Social desirability and Q sorts. *J. consult. Psychol.*, 1955, **19**, 462–463.
4. Ellis, A. The validity of personality questionnaires. *Psychol. Bull.*, 1946, **43**, 385–440.
5. Freeman, F. S. *Theory and practice of psychological testing*. New York: Holt, 1950.
6. Karson, S., & Freud, S. L. Predicting psychiatric diagnosis with the MMPI. *J. clin. Psychol.*, 1956, **12**, 376–378.
7. Kenny, D. T. The influence of social desirability on discrepancy measures between real self and ideal self. *J. consult. Psychol.*, 1956, **20**, 315–319.
8. Klett, C. J. The stability of the social desirability scale values in the Edwards Personal Preference Schedule. *J. consult. Psychol.*, 1957, **21**, 183–185.
9. Klett, C. J. The social desirability stereotype in a hospital population. *J. consult. Psychol.*, 1957, **21**, 419–421.
10. Lindquist, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
11. Rosen, E. Self-appraisal, personal desirability, and perceived social desirability of personality traits. *J. abnorm. soc. Psychol.*, 1956, **52**, 151–159.
12. Rubinstein, E. A., & Lorr, M. A comparison of terminators and remainers in outpatient psychotherapy. *J. clin. Psychol.*, 1956, **12**, 345–348.

Journal of Consulting Psychology
Vol. 22, No. 5, 1958

A Technique for Investigating the Relationship Between the Behavior Cues of the Examiner and the Verbal Behavior of the Patient[1, 2]

Leonard Krasner
Veterans Administration Hospital, Palo Alto, California

In a previous paper, the writer suggested that techniques of conditioning verbal behavior be used in a program of systematic investigation into the process of psychotherapy and offered a rationale for such a procedure (2). It was proposed that such a program be initiated by studying the relationship between behavioral cues indicating "attention" on the part of the examiner (independent variable) and changes in specified verbal behavior on the part of the patient (dependent variable). The purpose of this paper is to present an experimental technique that can be utilized in such an approach and its results with two patients.

A technique was devised to create an interpersonal situation which would have as many of the characteristics of the psychotherapeutic situation as possible. Further consideration was to control variables of experimenter's behavior and to quantify the subject's verbal behavior. It was hypothesized that systematic changes in the experimenter's behavior would result in specifiable changes in the subject's verbal behavior.

Method

Subjects. Two male, white, veterans, 30 years of age, with high school education, and diagnosed as schizophrenic reactions in remission, were asked to participate in a study to determine how people "tell stories."

Procedure. The equipment consisted of a standard timer, placed on the desk so that both *S* and *E* could see it, and a tape recorder. Prior to the first session, *S* was given the following instructions: "I want you to make up a story with at least four characters in it, a mother, a father, a child, and an animal. It can be any kind of a story. I want you to tell me the story for ten minutes. I'll start the timer now, and it will ring when the ten minutes have passed. All right, go ahead." Prior to each of the following 24 sessions, the same instructions were repeated with one change. Instead of "I'll start the timer now, and it will ring when ten minutes have passed," *E* said, "You can either continue your story from the last session or start a new one."

Each *S* participated in 25 ten-minute sessions, which were divided into five blocks of five sessions each. During the first block of five 10-minute sessions (nonreinforcement), *E* avoided looking at the *S*. *E* either looked down at the blank pad in front of him or looked out of the nearby window. During the second block of five sessions (reinforcement), *E* responded with a combination of behavior cues after each time the *S* made reference to the class of verbal behavior "mother." *E*'s combined behavior cue during the reinforcement condition consisted of looking at *S*, nodding his head, smiling, and emitting an "mmm-hmm" sound. The class of verbal behavior "mother" was defined as "all nouns and pronouns referring to the mother figure in the *S*'s story."

In the third and fifth block of five sessions

---

[1] From the Veterans Administration Hospital, Palo Alto, California. Research was conducted at the Veterans Administration Hospital, Lexington, Kentucky.

[2] A paper based on this study was presented at the 1957 American Psychological Association meetings.

Table 1

Number of "Mother," "Father," and Total "References" (Nouns and Pronouns), and Percentages of "Mother" to "Mother and Father" and to Total "References" for Successive Blocks of Five Ten-Minute Sessions

		Number references			Per cent references	
Subject and 10-minute sessions	Condition	"Mother"	"Father"	Total	"Mother"/ "Mother" and "Father"	"Mother"/ Total
Subject I						
1–5	nonreinf.	161	201	1306	45	12
6–10	reinf.	272	138	1315	66	21
11–15	nonreinf.	255	235	1212	52	21
16–20	reinf.	313	245	1222	56	26
21–25	nonreinf.	190	227	1197	46	16
Subject II						
1–5	nonreinf.	68	48	384	59	18
6–10	reinf.	90	51	319	64	28
11–15	nonreinf.	18	109	261	14	7
16–20	reinf.	129	114	519	53	25
21–25	nonreinf.	100	153	552	40	18

the nonreinforcement condition, as in the first block of five sessions, was repeated. The fourth block of five sessions repeated the reinforcement condition of the second block of five sessions.

After completion of the 25th session, the *S*s were interviewed to determine their degree of awareness of the conditioning procedure. This interview consisted of a dozen questions such as, "What do you think was the purpose of these sessions," "Did you notice any changes in the stories as you went on," "Do you think that my sitting here in any way influenced your stories?"

Results

The total number of "references" made by *S* to each of the major characters in the stories —"mother," "father," "child," and "animal" —were summarized for each individual session. To be scored as a "reference," the verbalization had to be a noun or pronoun referring to one of the four major characters. The ratios of "mother references" to "mother–father references" and to total "references" were computed for each of the 25 sessions.

These two ratios were divided into those verbalizations emitted under reinforcement conditions and those verbalizations emitted under nonreinforcement conditions. The Mann-Whitney *U* test (5, pp. 116–126) for the significance of differences of "references" under the conditions of reinforcement and nonreinforcement for these two ratios was computed for both subjects. All four of the ratios were in the predicted direction, and three of the four, both ratios of S_{I}, and the "Mother"/total ratio for S_{II}, were significant beyond the .05 level.

An examination of these two sets of ratios for successive blocks of five sessions indicates that the ratios of the reinforced class of behavior increased under reinforcement conditions and decreased under nonreinforcement, increased again during the succeeding reinforcement series, and decreased again during the final block of nonreinforcement sessions. These systematic variations were also true of the absolute number of "references" as well as the percentages. Table 1 presents both percentages and numbers of "references" for successive blocks of five 10-minute sessions.

Discussion

This technique of verbal conditioning enables the testing of hypotheses concerning what happens to various classes of verbal behavior under selected experimental conditions. The hypothesis in the present study that the reinforced class of verbalization varies systematically with the application of *E*'s behavior cues was confirmed with two patients. It is of interest to note what happened to the other categories of verbal behavior which appeared in the stories, i.e., "father," "child," "animal." All three of these categories varied sharply from session to session, but in no systematic manner.

In a review of "verbal conditioning" studies with their implications for psychotherapy (3), the writer divided the techniques of conditioning verbal behavior into four general categories of experimental situations: saying plural nouns; completing sentences; "storytelling" and interview situations; and test-like situations. The technique used here falls in the "storytelling" category and is similar to the techniques reported in studies by Ball (1) and Mock (4).

Summary

A storytelling technique was used to study the relationship between examiner behavior cues and patient's verbal behavior. The results indicate that changes in a preselected class of verbal behavior varied as a function of the systematic application of behavior cues by the examiner. It is concluded that this experimental procedure permits the systematic isolation and study of important variables of the interpersonal process basic to psychotherapy.

Received October 8, 1957.

References

1. Ball, R. S. Reinforcement conditioning of verbal behavior by verbal and non-verbal stimuli in a situation resembling a clinical interview Unpublished doctoral dissertation, Indiana Univer., 1952.
2. Krasner, L. The use of generalized reinforcers in psychotherapy research. *Psychol. Rep.*, 1955, **1**, 19–25.
3. Krasner, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55**, 148–170.
4. Mock, J. F. The influence of verbal and behavioral cues of a listener on the verbal productions of the speaker. Unpublished doctoral dissertation, Univer. of Kentucky, 1957.
5. Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Journal of Consulting Psychology
Vol. 22, No. 5, 1958

Q—L Variability, MMPI Responses, a College Males

William D. Altus
University of California, Santa Barbara Coll

One of the fundamental cleavages in intelligence as it is commonly tested is that obtaining between the ability to manipulate verbal concepts effectively and the ability to reason with speed and efficiency in terms of quantities. This broad dichotomy of general aptitude is recognized by the College Entrance Examination Board in its Scholastic Aptitude Test (SAT) which is comprised of two tests, the Mathematics Aptitude Test (MAT) and the Verbal Aptitude Test (VAT). In like manner, the American Council on Education Psychological Examination in its decades of existence had its separate quantitative and linguistic scores. The Army General Classification Test of World War II sampled these same two abilities, linguistic and quantitative; a third ability measured by the AGCT was a spatial factor. It would appear, therefore, that both theory and practice in the construction of measures of general aptitude have placed the stamp of approval on the functional independence of these two large sectors of one's intellectual abilities, i.e., dealing effectively and precisely with quantities on the one hand and with the less precise, often affectively toned verbal constructs on the other.

It would seem plausible to expect that individual variability along such culturally significant continua as quantitative and linguistic abilities might be reflected in one's style of living and in his various attitudinal frames. F. L. Wells (6), for instance, feels that these aspects of intelligence are significantly related to personality attributes. His basic contention appears to be that a deficiency in quantitative ability, say, is a function of a basic attitude: An individual's being repelled by the rigidity and precision of a numerical quantity, being attracted by the adumbrative and connotative nuances of the adjective and the verbal tag.

The verbalist p F such a statement as
"That day we t a long and dusty dis-
tance," to the r osaic and exact, "We
spent ten hours T 14.2 miles that day."
To the verbalis latter statement limits
and confines, le nothing to the imagina-
tion, is obtrusiv its specificity and its
fractions. Freud Fa he a good example of
the verbalist. H distrusted sta-
tistics; he even Fa work out
his traveling sch oad, since
the thicket of n False nary time-
table was appar ability to
handle. Jones True (2) tement of
Freud's concerni qu-linguistic
dichotomy which False nteresting,
"I have very res or talents.
None at all for False s; nothing
for mathematics; thing quanti-
tative."

Wells appears True hat what he calls
verbalism is no r Fals entity than is a neu-
rosis, but smacks Falsd, of a temperamental
characteristic rat n an intellectual one.
In terms of gene ustment, he feels, the
verbalist and the True tatively oriented per-
son should be r False coequal, though the
verbalist may be hat more in accord
with the norms o al prestige, which re-
ward precision a ty of diction in both
speaking and wri ells seems to approve
completely of a nt of Adler's to the
effect that the tively oriented indi-
vidual is more s cient, while the ver-
balist tends to s urity in social rela-
tions. "This," W ntinues (6, p. 72),
"ties in with the social tendency ob-
served generally i erbal Facility' group,
as well as the spe ed for 'belonging' in
this particular man
In 1946, Munroe (res ublished some inter-

esting findings sse certain differ-
ences she found no Rorschach be-
tween students nfrence who were
quantitatively c aneasured by the
Q and L fractio a) in comparison
with those who went in the op-
posite direction. conclusions re-
lating to the sp nces found were
that the high Q relatively literal
in their constru e, while the high
L women used pective approach.
Pemberton (4), male executives,
was able to sh *nificant relations*
between L and ls on the ACE
and certain sco uder Preference,
Allport-Vernon lues, Thurstone
Temperament lford-Martin In-
ventory of Fact For example, the
men whose L s gma above their
Q scores had igher scores on
the Kuder Liter urstone's Reflec-
tive factor, the n Social.
In a previous Altus (1) found
some significant een Q—L differ-
entials on the A swers of 200 col-
lege women to th Multiphasic Per-
sonality Invent l, he found that
the *Q-higher-th* oman was prim,
conventional, ious, and some-
what resentful with the college
woman whose to go in the op-
posite directic . A scale of 43
items was der MMPI which on
cross-validatic of .25, which is
just short of a e .01 level, since
the *N* involve e recently, Spilka
and Kimble (Altus' 43 MMPI
items with a p women students
at Washburn und an *r* of .22
with Q—L di e ACE, thus con-
firming in a the validity of
the items in pears, therefore,
from these p ted studies that
variation in the quantitative-
verbalist con ed in certain per-
sonality fact be specified in
crude degree en, at least.

tudy

Since college been previously
investigated, it w d to confine the present research to college men, using the same psychometric variables, i.e., the MMPI and the differentials in Q and L on the ACE. Two hypotheses were set up: (*a*) that the *Q-higher-than-L* college male would be more masculine (have a lower *Mf* score on the MMPI) than the *L-higher-than-Q* college male; (*b*) that the *L-higher-than-Q* men should appear somewhat more sophisticated and mature on certain of their answers. The reasoning behind the hypothesis of greater masculinity of the high Q males is that men, generally, tend to do better on Q tests than do women; hence, it would appear that men who excel other men in this sex-biassed variable should show more of those attitudes culturally associated with masculinity. The basis for the second hypothesis is derived from local findings on college women: the more verbal women were more mature and sophisticated in the answers which discriminated them from the higher Q women on the MMPI.

The raw ACE Q and L scores for 200 college males were converted into standard units with means of 50 and sigmas of 10. Two groups of 100 with comparable Q and L scores were formed. Each of these groups of 100 was broken into quartiles of 25, the basis for the separation being the direction of the discrepancy scores between the Q and L variables. All the validating and clinical scales of the MMPI were then tested for significant mean differences. In addition, the individual items of the MMPI were analyzed for relations to the Q—L dichotomy. Only those items which showed consistent differences in the same direction for both groups of 100 males were retained for study. This precaution was necessary because of the probably low reliability of individual T-F items in the MMPI; using the entire 200 as a single group with four quartiles for analysis would cause the inclusion, by chance alone, of a sizable, if indeterminate, number of items from the 566 items which comprise the MMPI.

Results

Two of the MMPI scales showed significant mean differences between upper and lower quartiles when the entire 200 college men were considered as a single group and ranked on

Table 1

The Significance of the Differences Between the Answers of 100 High Q and 100 High L College Males to Selected Items of the MMPI

Gp. MMPI No.	Level of confidence	Marking characteristic of Q males	MMPI items
225	.001	False	I gossip a little at times.
244	.001	True	My way of doing things is apt to be misunderstood by others.
295	.001	False	I liked "Alice in Wonderland" by Lewis Carroll.
384	.001	True	I feel unable to tell anyone all about myself.
504	.001	True	I do not try to cover up my poor opinions or pity of a person so that he won't know how I feel.
564	.001	True	I am apt to pass up something I want to do when others feel that it isn't worth doing.
126	.01	False	I like dramatics.
127	.01	False	I know who is responsible for most of my troubles.
140	.01	False	I like to cook.
204	.01	False	I would like to be a journalist.
237	.01	False	My relatives are nearly all in sympathy with me.
366	.01	True	Even when I am with people I feel lonely much of the time.
443	.01	True	I am apt to pass up something that I want to do because others feel that I am not going about it in the right way.
448	.01	True	I am bothered by people outside, on street-cars, in stores, etc. watching me.
45	.05	False	I do not always tell the truth.
69	.05	True	I am very strongly attracted by members of my own sex.
78	.05	False	I like poetry.
82	.05	True	I am easily downed in an argument.
325	.05	False	The things that some of my family have done have frightened me.
404	.05	True	People have often misunderstood my intentions when I was trying to put them right and be helpful.
488	.05	False	I pray several times a week.
520	.05	False	I strongly defend my own opinions as a rule.
95	.10	False	I go to church almost every week.
180	.10	True	I find it hard to make talk when I meet new people.
188	.10	False	I can read a long while without tiring my eyes.
255	.10	False	Sometimes at elections I vote for men about whom I knew very little.
321	.10	True	I am easily embarrassed.
428	.10	False	I like to read newspaper editorials.
64	NS	False	I sometimes keep on at a thing until others lose their patience with me.
258	NS	True	I believe there is a God.
277	NS	False	At times I have been so entertained by the cleverness of a crook that I have hoped he would get by with it.

the continuum of Q—L discrepancy on the ACE. For both of the separate groups of 100 men, the quartile containing the men with the highest *Q-higher-than-L* discrepancy score showed higher mean values on the MMPI Lie scale than did the quartile for the men at the opposite pole, where the *L-higher-than-Q* males were found. The mean difference was significant at the .01 level. This finding would tend to bear out in a gross way the hypothesis that the *L-higher-than-Q* males tend to be somewhat more sophisticated or less naive than the *Q-higher-than-L* college male.

Also in line with the original hypotheses, the *Mf* scale of the MMPI showed a highly significant mean difference between the high and low quartiles of the total group of 200 in Q—L discrepancy scores. The mean difference is significant at the .001 level of confidence (t of 4.30). The college male whose abilities are relatively higher in dealing with quantities does have more masculine attitudes, insofar as these are validly measured by the *Mf* scale of the MMPI. None of the other MMPI scales, either clinical or validating, showed any mean difference that was significant even at the .05 level, corroborating once more Wells' original contention and Altus' and Munroe's findings that the Q—L discrepancies on the ACE are not related to general adjustment for college women.

Some 22 MMPI items which differentiate from the .05 to the .001 level between the Q- versus the L-oriented college man are given in Table 1; an additional six which differentiate at the .10 level are also listed, as well as three whose nominal significance is less than .10. It will be remembered that two groups of 100, each of which was broken into four quartiles for purposes of analysis, were used in sifting the 566 MMPI items. Only those items which inspection indicated to vary rather markedly and in the same direction for the two groups were picked out to test for levels of confidence. Before confidence levels were tested, all 100 males in both groups who were above the average for the total group in Q score were compared with the 100 whose abilities ran to the verbal; in other words, the *entire* group of 200 males was used in these comparisons—a technique which tends to attenuate the differences which emerge when only extreme groups are employed.

Six of the items in Table 1, Nos. 69, 78, 126, 140, *204*, and *295*, are found on the *Mf* scale. On five of these items the marking shows a correlation between masculinity and *Q-higher-than-L* tendency; on the sixth item, "I am very strongly attracted by members of my own sex," the marking is in the opposite direction, that is, the *Q-higher-than-L* males say Yes to this item more frequently (.05 l.c.). The *Q-higher-than-L* female answered the same way (1) and at the .001 level of confidence. It is difficult to rationalize this item. It could mean that homosexual attitudes are more prevalent among high Q groups, male and female; it might also mean that they are less well socialized, rather timid, and at ease only with their own sex. However that may be, the dislike for poetry, cooking, journalism, dramatics, and Lewis Carroll's *Alice in Wonderland* stands out as a typical anti-aesthetic attitude common among the more "masculine" males.

Three items from the Lie scale show up: 45, *225*, 255. The high Q male is more frequently certain he always tells the truth, that he doesn't gossip, and that he knows something about all those men he votes for at election time. Naivete and lack of insight seem to be the plausible explanation rather than an attempt at dissimulation, since dissimulation would be reflected in attempts to "fake good" on the clinical scales which patently is not the case. It may, of course, actually be true that the high Q man doesn't gossip as often as his verbalist brother; just possibly he doesn't converse as frequently and perhaps he isn't as interested in people—this latter observation, it should be noted, would fit in with Wells' and Adler's contentions.

The attitude of the high Q males toward religious matters (*95*, *258*, *488*) is somewhat ambiguous: They pray less frequently than the more verbal college student, and they go to church less frequently. Although the answer is not statistically significant, they more frequently are sure there is a god (258): Of the 100 *Q-higher-than-L* men, 88 said Yes to this item, while only 80 of the *L-higher-than-Q* men answered in this fashion. In Altus' study (1) of college women, the high Q women were definitely more religiose, but it must be admitted that the evidence for college men is equivocal. In both instances, however, the high Q student is more orthodox in belief even if he isn't necessarily in his religious practice.

The dislike for things verbal stands out in several of the high Q answers: 78, 126, 188, 204, 295, 428. They don't like poetry, they don't like dramatics, they "tire" their eyes if they read a lot, they wouldn't like to be a journalist, they didn't like Carroll's *Alice*, they don't like to read editorials in the newspapers. This finding strikingly corroborates the data for college women (1). The high Q person simply does not like to read to the same extent as does the verbalist. Which is cause, however, and which is effect is unclear, even though the facts of the matter are patent.

The high Q person is not so dominant or sure of himself in certain matters: 82, 443, 520, 564. He's easily downed in an argument, he passes things up when others criticize, he doesn't defend his opinions strongly as a rule. He is therefore less dominant in interpersonal relationships; for this reason we should expect to find, perhaps, that most salesmen and almost all politicians would be on the verbalist side.

Constriction and difficulty in interpersonal relations is also evident in many other answers: 82, 180, 237, 321, 366, 384, 404, 443, 504, 520, 564. Other people outtalk him, he finds it difficult to make conversation with new people, his relatives are not in sympathy with him, he is easily embarrassed, he feels lonely even when with people, he can't unburden himself freely to others when talking of himself, he passes up things when others criticize, he doesn't mask his attitude toward others when he should. A rather gauche, retiring, somewhat socially inept person is here adumbrated. A number of these items also characterized the high Q woman (1) but not to the same degree as is noted here for the college male.

Naiveté, dislike of the printed page, difficulty in social relationships, greater masculinity (put more exactly, antipathy to the aesthetic)—these are the general characteristics which the MMPI items show to be true to the *Q-higher-than-L* college male when he is compared with his verbalist brother.

Discussion

The assumption which underlay the present research was that differences in quantitative and linguistic abilities among college males, as measured on the ACE, should be reflected in attitudinal frames as measured by the MMPI. It was hypothesized that the *Q-higher-than-L* male should tell more "lies" on the MMPI and appear more masculine. Both hypotheses were born out at statistically significant levels when the *L* and *Mf* scales were checked for the two groups of men.

The individual items of the MMPI which differentiated high Q and high L groups corroborated previous research (1, 4, 5) which showed an antipathy toward reading and similar verbal matters and occupations for the quantitatively oriented person. Both male and female high Q scorers also appear to be less forward, aggressive, and sure of themselves in social relationships than the student with more verbalist propensities. The data lend some plausibility to a hypothesis that social dominance and leadership may show more than a chance linkage to *L-higher-than-Q* type of abilities. It would appear that the politician, who lives primarily in a welter of words, should be a verbalist far more frequently than the 50/50 probability of the flip of a coin turning up heads. These tentative hypotheses lend a bit of piquancy when one speculates about Freud and his marked deficiency in matters quantitative. Had he been more adept at manipulating numbers, perhaps he would have remained in laboratory research; had this obtained, there would have been no private practice of medicine, no neurotics, no dream theories, no psychoanalysis. But this is pure speculation, interesting perhaps but in the view of history completely idle.

The contention of Wells (6) and Munroe (3) that general adjustment is not related to variation in quantitative and verbal abilities appears to be corroborated by the present study and by the one which preceded it (1). In summary, the one whose gifts are primarily verbal tends to be more sophisticated, self-insightful, socially dominant, literary, and less orthodox in religious matters; at least that is the way college students of both sexes tend to evaluate themselves on the MMPI and, to the extent that their answers are truthful and valid, it should show in actual behavior outside the narrow limits of the testing situation.

Received June 16, 1958.
Early Publication.

References

1. Altus, W. D. Personality correlates of Q—L variability on the ACE. *J. consult. Psychol.*, 1952, **16**, 284–291.
2. Jones, E. *Sigmund Freud, life and work*. Vol. 2. London: Hogarth, 1955.
3. Munroe, Ruth L. Rorschach findings on college students showing different constellations of subscores on the ACE. *J. consult. Psychol.*, 1946, **10**, 301–316.
4. Pemberton, C. L. Personality inventory data related to ACE subscores. *J. consult. Psychol.*, 1951, **15**, 160–162.
5. Spilka, B., & Kimble, Gloria. Personality correlates of Q—L differentials on the ACE. *J. consult. Psychol.*, 1958, **22**, 142.
6. Wells, F. L. Verbal excess over quantitation: Two case studies. *J. Psychol.*, 1946, **23**, 65–82.

Journal of Consulting Psychology
Vol. 22, No. 5, 1958

Some Scale and Test Correlates of a Personality Rigidity Scale[1, 2]

John M. Rehfisch
Sonoma State Hospital, Eldridge, California

This is a report of the correlations between a true–false inventory scale for personality rigidity (*Ri*) and numerous other objective psychological measures. An earlier paper (5) described the theory and derivation of *Ri*. Briefly, the scale was constructed by selecting items which differentiated between designated samples of rigid and flexible subjects. The criterion employed was ratings of rigidity by staff assessors from the University of California's Institute of Personality Assessment and Research (IPAR). A cross-validating appraisal of a preliminary version of *Ri*, in terms of staff descriptions of high and low scorers, indicated that the scale measures essential attributes of the conceptualized rigidity dimension. High scorers, as compared to lows, were described generally as constricted, inhibited, anxious, guilt prone, conservative, socially introverted, and inflexible in their social roles. Low scorers, by contrast, tended to be seen as adaptable, spontaneous, original, fluent in thought and speech, curious, clear-thinking, assertive, and self-indulgent.

The present study was designed to further explicate *Ri* by defining its linear relations to other psychological scales and tests.

Procedures and Results

Correlation coefficients were computed between *Ri* and: (*a*) the standard MMPI scales, (*b*) three special MMPI scales (1, 4, 7), (*c*) the standard scales of the California Psychological Inventory (CPI) (2), (*d*) the Terman Concept Mastery Test (6), (*e*) the Idea Classification Test,[3] and (*f*) tests from the Guilford Creativity battery (3). Two samples of AF captains (60 in each sample), separate from the assessed samples used in the derivation and initial validation of the scale, were employed as subjects. Their scores on the above listed measures were obtained from results of an IPAR paper-and-pencil testing program, which had been administered to the officers at their home bases.

Correlations were computed also between *Ri* and the UCPOS Ethnocentrism and Fascism scales. The single sample of 100 assessed AF captains used in obtaining the latter two correlations had been included among the groups employed in deriving *Ri*: this sample was the only one for which scores on both *Ri* and the UCPOS scales were available.

All the findings are listed in Tables 1, 2, 3, and 4.

Prominent among the major correlates are various indices of social behavior and of dominance and leadership ability. High scorers

---

[1] This research is supported in part by the United States Air Force under Contract No. AF 18 (600)-8, monitored by Technical Director, Detachment No. 7 (Officer Education Research Laboratory), Air Force Personnel and Training Research Center, Maxwell Air Force Base, Alabama. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government. Personal views or opinions expressed or implied in this publication are not to be construed as necessarily carrying the official sanction of the Department of the Air Force or of the Air Research and Development Command.

[2] The author is very much indebted to Harrison Gough for his advice and encouragement and for his critical reading of a preliminary version of the manuscript; and to Donald MacKinnon, Director of the Institute of Personality Assessment and Research, for many valuable suggestions and for permission to use the data and facilities of the Institute.

[3] Acknowledgment is made to Educational Testing Service for making this test available prior to its publication.

Table 1

Correlations of *Ri* with Scales of the MMPI for Two Samples of AF Captains

Scale	r ($N = 60$)	r ($N = 60$)
L	−.04	−.29*
F	.17	.17
K	−.57**	−.59**
Hs	−.10	.05
D	.32*	.43**
Hy	−.26*	−.25
Pd	−.17	−.10
Mf	−.16	.16
Pa	.00	−.16
Pt	.18	.41**
Sc	−.08	−.04
Ma	−.21	.03
Si	.64**	.73**
A (1st factor: anxiety)	.48**	.57**
Es (ego-strength)	−.44**	−.59**
Lp (leadership)	−.66**	−.65**

* Significant at the .05 level.
** Significant at the .01 level.

on *Ri*, as opposed to lows, would appear to be socially introverted (*Si*, *Sy*), submissive (*Do*), and relatively deficient in personality attributes associated with social presence (*Sp*) and with leadership ability (*Lp*).

Table 2

Correlations of *Ri* with Scales of the CPI for Two Samples of AF Captains

Scale	r ($N=60$)	r ($N=60$)
Do (dominance)	−.53**	−.59**
Cs (capacity for status)	−.55**	−.55**
Sy (sociability)	−.70**	−.72**
Sp (social presence)	−.75**	−.74**
Sa (self-acceptance)	−.29*	−.41**
Wb (sense of well-being)	−.38**	−.49**
Re (responsibility)	−.16	−.34**
So (socialization)	−.08	.10
Sc (self-control)	.06	.19
To (tolerance)	−.55**	−.61**
Gi (good impression)	−.32*	−.35**
Cm (communality)	−.03	−.27*
Ac (achievement via conformance)	−.52**	−.50**
Ai (achievement via independence)	−.50**	−.23
Ie (intellectual efficiency)	−.64**	−.62**
Py (psychological-mindedness)	−.24	−.39**
Fx (flexibility)	−.33*	−.31*
Fe (femininity)	.29*	.21

* Significant at the .05 level.
** Significant at the .01 level.

Table 3

Correlations of *Ri* with Tests of Cognitive Functioning for Two Samples of AF Captains

Test	r ($N=60$)	r ($N=60$)
Intelligence tests:		
Terman Concept Mastery	−.33*	−.16
Idea Classification	−.21	−.16
Creativity tests:[a]		
Unusual Uses	−.34**	−.28*
Gestalt Transformation	−.23	−.24

* Significant at the .05 level.
** Significant at the .01 level.
[a] Correlations were computed, in one of the above samples, between *Ri* and other tests from the creativity battery, including Plot Titles, Match Problems, Consequences, and Controlled Association. These correlations were all approximately zero.

A second cluster of correlates suggests that high *Ri* persons are likely to be anxious and depressed (*A*, *D*), self-dissatisfied, and emphatic in the expression of their complaints (*Sa*, *K*, *Wb*).

There are indications also that high scorers tend to be less intellectually efficient (*Ie*), less interested in and motivated toward scholastic achievement (*Ac*, *Ai*),[4] less original (Unusual Uses), and slightly less intelligent than low scorers.

The correlations of *Ri* with the Tolerance, Fascism, and Ethnocentrism scales accord with the findings from various other studies which reflect an association between rigidity and prejudice.

The *Es* scale, which is negatively related to *Ri*, was derived by contrasting the item responses of patients who did and patients who

Table 4

Correlations of *Ri* with Scales of the UCPOS for a Sample of AF Captains

Scale	r ($N=100$)
F (fascism)	.25*
E (ethnocentrism)	.23*

* Significant at the .05 level.

---

[4] The *Ac* and *Ai* scales were derived by selecting items capable of discriminating between high and low achievers at the high school (*Ac*) and college (*Ai*) levels.

did not benefit from psychotherapy (1). High *Es* scores are predictive of a favorable therapeutic outcome. The present implication of a negative relation between rigidity and capacity for improvement in psychotherapy appears reasonable; for it is quite probable that the modification in patient personality, seemingly involved in a beneficial therapeutic process, is facilitated by psychological flexibility.

The negative correlations with the Flexibility scale were naturally anticipated. That the correlations are not larger is probably a consequence of essential dissimilarities between the two scales due to differences in the methods used in their construction. The *Fx* scale was developed in a rational a priori manner (2), whereas *Ri* was derived by empirical analysis.

Finally, there are indications (*Cs*) that qualities assessed by *Ri* are negatively related to capacity for high social status.

Summary and Conclusions

The test and scale correlates of a scale for personality rigidity indicate a tendency for high scorers on *Ri*, as contrasted to lows, to be: (*a*) socially introverted and lacking in social presence (defined as poise, spontaneity, and self-confidence [2]); (*b*) submissive and low in leadership qualities; (*c*) anxious and self-disparaging; and (*d*) unoriginal and relatively deficient in cognitive and motivational factors associated with intellectual competence and achievement. These findings are generally congruent with other validating data (5), including the assessor descriptions of high and low scorers, obtained among subjects separate from those employed in the present study. The qualities listed above may therefore be considered fairly stable *Ri* correlates for the kinds of subjects used in defining the scale, i.e., adult males (university students [5] and AF captains) who were, as a group, above the average in intelligence and education.

There were indications also, among the correlating scales, of a positive association between *Ri* and prejudice. Such a relationship had not been specifically indicated by the assessor descriptions; prejudice may, accordingly, be a somewhat less salient and/or reliable correlate of the *Ri* scale than the other qualities listed above.

Received September 3, 1957

References

1. Baron, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, 17, 327–333.
2. Gough, H. G. *The California Psychological Inventory*. Palo Alto, Calif.: Consulting Psychologists Press, 1956.
3. Guilford, J. P., Wilson, R. C., Christensen, P. R., & Lewis, D. J. A factor-analytic study of creative thinking: I. Hypotheses and descriptions of tests. *Rep. Psychol. Lab., Univer. Southern Calif.*, 1951, No. 3.
4. Oettel, A. M. Leadership: A psychological study. Unpublished doctoral dissertation, Univer. of California, 1953.
5. Rehfisch, J. M. A scale for personality rigidity. *J. consult. Psychol.*, 1958, 22, 11–15.
6. Terman, L. M., & Oden, M. H. *The gifted child grows up*. Palo Alto, Calif.: Stanford Univer. Press, 1947.
7. Welsh, G. S. Factor dimensions A and R. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956. Pp. 264–281.

Journal of Consulting Psychology
Vol. 22, No. 5, 1958

The Effect of Cultural Factors on the Relationship Between Personality, Intelligence, and Ethnocentric Attitudes[1]

Aron Wolfe Siegman
Bar-Ilan University, Israel

Many investigations have been undertaken in order to determine some of the variables which are associated with ethnocentric attitudes. Frenkel-Brunswik and her associates have reported fairly high correlations between ethnocentric attitudes and authoritarian ideology of the fascistic variety (1). Some authors have suggested that ethnocentric attitudes are associated with personality maladjustment (6, 7, 9) and lower intelligence (6, 7, 12). Very little attention, however, was paid to the problem whether these obtained relationships may not be a function of the degree to which the culture sanctions ethnocentric attitudes. The major purpose of the present study was to investigate the relationship between authoritarian ideology, personality maladjustment, and intelligence on the one hand, and anti-Negro attitudes on the other hand, in a group of college students who were born, raised, and at the time of this study resided in the South.

Method

Subjects. The *Ss* were 41 University of North Carolina undergraduates, all of whom were born, raised, and at the time of this study resided in the South.

Measures. All *Ss* were administered the six items from the E scale which pertain to Negroes (1, p. 142), and a 29-item F scale (1, pp. 255–257). The Taylor MAS (13), which was found to correlate .92 with a general neuroticism scale (5), was administered as an index of *Ss*' adjustment level. All *Ss* were also administered Gough's *Pr* scale (7), which consists of personality oriented MMPI items which discriminated between high and low scorers on an anti-Semitism scale. There is evidence to suggest that a high *Pr* scale score is associated with personality maladjustment (7, 9).

It has been demonstrated that vocabulary test performance correlates very highly with performance on an intelligence test which samples a wide variety of intellectual abilities (14). In the present study, the vocabulary test of the Shipley-Hartford (S-H) Retreat Scale (10) was used to determine *Ss*' intelligence level. In order to raise the ceiling of the test, a three-minute time limit was imposed.

In a recent study which investigated the relationship between *Ss*' F scale and Einstellung test scores, it was found that it was not so much the content of the F scale items, but rather *Ss*' tendency to "acquiesce" and answer "True," irrespective of content, which was responsible for the positive correlation between *Ss*' F scale and Einstellung test scores (8). This finding raises the problem whether the same explanation may not be true for the frequently noted negative correlation between *Ss*' F scale and intelligence test scores (1, 3, 6, 11). Consequently, all *Ss* in this study were also administered a reversed F scale (2).

In order to minimize defensiveness, *Ss* were assigned numbers and requested not to indicate their names on the attitude and personality questionnaires.

Results and Discussion

Table 1 lists *Ss*' mean and standard deviation scores on the various tests. Table 2 indicates that no significant correlations were obtained between *Ss*' E scale scores on the one hand, and their Taylor MAS, Gough *Pr*, and S-H vocabulary test scores on the other hand.

---

[1] This investigation was conducted while the author was at the University of North Carolina. The study was supported in part by a grant from the Behavioral Science Research Fund.

This finding is clearly consistent with the hypothesis that neither neuroticism nor intelligence is significant sources of variance in culturally accepted ethnocentric attitudes.

Table 2 indicates that a significant positive correlation was obtained between *S*s' E scale and F scale scores. The correlation, however, was significantly lower than the correlations which have been obtained by the authors of *The Authoritarian Personality* (1, p. 263) for non-Southern groups (average $r = .77$). This finding suggests that the magnitude of the correlation between a particular set of ethnocentric attitudes and authoritarian ideology, or authoritarianism as a personality variable (1), decreases with increased cultural approval of such ethnocentric attitudes.

Table 2 indicates that a significant negative correlation was obtained between *S*s' F scale (but not reversed F scale) and S-H vocabulary test scores. Significant positive correlations were obtained between *S*s' F scale scores on the one hand, and their Taylor MAS and Gough *Pr* scale scores on the other hand. These results are consistent with the hypothesis that intelligence and neuroticism are significant sources of variance in authoritarian ideology. The finding, however, that neither neuroticism nor intelligence was a significant source of variance in culturally approved ethnocentric attitudes, certainly suggests the possibility that the contribution of neuroticism and intelligence to authoritarian ideology too may decrease with increased cultural approval of such ideology. In fact, the correlation between *S*s' Taylor MAS and F scale scores in the present study was significantly lower than the correlation which was obtained by Davids (3) in a group of Harvard students (.69) whose mean score per item on the F scale (2.97) was lower than that of the present group (3.43).[2]

Table 1

*S*s' Mean and *SD* Scores on the Various Psychological Tests

Tests	Mean	*SD*
F scale	99.51	15.37
E scale	19.73	6.60
Taylor MAS	15.15	7.80
Gough *Pr* scale	10.51	5.01
S-H Vocabulary MA	16.42	1.15

Table 2

Correlations Between the E and F Scales and Other Psychological Tests

Tests	S-H Vocabulary	Taylor MAS	Gough *Pr*	E scale	F scale
E scale	−.03	.16	.06		
F scale	−.38*	.31*	.57**	.41*	
F scale (reversed)	−.14	.03	.08		−.001

* Significant at the .05 level.
** Significant at the .01 level.

Summary

No significant correlation was obtained, in a group of Southern undergraduates, between *S*s' performance on an anti-Negro bias (E) scale and *S*s' Taylor MAS, Gough *Pr* and Shipley-Hartford vocabulary scores. Furthermore, the correlation between *S*s' E and F scale scores was significantly lower than in non-Southern groups. These findings are consistent with the general hypothesis that the negative correlates of ethnocentric attitudes tend to decrease as the culture countenances these attitudes.

Received September 24, 1957.

References

1. Adorno, T. W., Frenkel-Brunswik, Else, Levinson, D. J., & Sanford, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. Bass, B. M. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 616–623.
3. Davids, A. Some personality and intellectual correlates of intolerance of ambiguity. *J. abnorm. soc. Psychol.*, 1955, **51**, 415–420.
4. Davids, A., & Eriksen, C. W. Some social and cultural factors determining relations between authoritarianism and measures of neuroticism. *J. consult. Psychol.*, 1957, **21**, 155–159.

---

[2] In a study which has appeared since the preparation of this paper (4), the authors arrived at a conclusion which is similar to the one suggested here. Data presented in that study made it possible to evaluate the significance of the difference between the mean F scale score of the Harvard group and the mean F scale score of the University of North Carolina group. The results were: $t = 2.41$, $p < .05$.

5. Eysenck, H. J. Reminiscence, drive, and personality theory. *J. abnorm. soc. Psychol.*, 1956, 53, 328–333.
6. Gough, H. G. Studies of social intolerance: I. Some psychological and sociological correlates of anti-Semitism. *J. soc. Psychol.*, 1951, 33, 237–246.
7. Gough, H. G. Studies of social intolerance: II. A personality scale for anti-Semitism. *J. soc. Psychol.*, 1951, 33, 247–255.
8. Jackson, D. N., Messick, S. J., & Solley, C. M. How "rigid" is the "authoritarian"? *J. abnorm. soc. Psychol.*, 1957, 54, 137–140.
9. Jensen, A. R. Authoritarian attitudes and personality maladjustment. *J. abnorm. soc. Psychol.*, 1957, 54, 303–311.
10. Shipley, W. C. *Shipley-Institute of Living Scale for measuring impairment.* Hartford, Conn.: Institute of Living, 1946.
11. Siegman, A. W. Some factors associated with authoritarian attitudes in children. Paper read at American Psychological Association, Chicago, September, 1956.
12. Siegman, A. W. Some variables associated with anti-Negro bias in children. Paper read at Eastern Psychological Association, New York, April, 1957.
13. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, 48, 285–290.
14. Wechsler, D. *The measurement of adult intelligence.* (3rd ed.) Baltimore: Williams & Wilkins, 1944.

Journal of Consulting Psychology
Vol. 22, No. 5, 1958

A Note on the Behavioral Validity of the Gordon Personal Profile

Warren W. Willingham, Paul Nelson, and William O'Connor[1]
U. S. Naval School of Aviation Medicine

Whereas personality scales are designed to indicate behavioral tendencies, it is generally difficult to find consistent behavioral criteria upon which to validate the scales. The present study takes advantage of such behavioral criteria extant in the Naval Air Training Program. While it is understood that a student may withdraw from flight training at any time, there is considerable pressure from peers and instructors to "stick it out." Some students drop for good reason, but in most cases the action is no clearer than a vaguely expressed "distaste for the situation." The Gordon Personal Profile has a scale termed *Responsibility* which is defined as follows: "Individuals who are unable to stick to tasks that do not interest them, and in the extreme, who tend to be flighty or irresponsible, usually make low scores on this scale" (1). The primary purpose of this study was to validate this scale against the criterion of voluntary withdrawal. A matter of secondary interest was the extent to which the *Emotional Stability* scale of the Profile was related to flight failure. A negative relation would be expected since anxiety or tenseness often hampers flight performance.

In the fourteenth week of preflight training, 1039 cadets completed the Gordon Personal Profile. Six months later, 95 had flight failed, and 117 had withdrawn voluntarily. The remaining 887 cadets were categorized as "successful." Table 1 shows the point-biserial correlation between each scale and each type of attrition. The data show consistent zero relationships. The frequency distributions of Total Score and the two subscales in question were examined in search of a discriminating cutoff which might be used in lieu of linear correlation statistics. From this analysis it was evident that the very low scoring students did not fail or withdraw at a higher rate.

Table 1

Means, Standard Deviations, and Validity Coefficients on the Gordon Personal Profile

Scale	*M*	*SD*	Validity coefficient: Withdraw	Validity coefficient: Failure
Ascendancy	5.2	5.3	.03	.00
Responsibility	7.2	5.0	.03	−.03
Emotional Stability	8.0	5.5	.04	−.04
Sociability	5.0	5.6	.00	−.01
Total Score	25.5	12.5	.04	−.03

In summary, the scales of the Gordon Personal Profile were not related to the behavioral criteria employed here. Recognizably, the criteria are complex, but at least some positive relationship should have been found. These data are inconsistent with the contention that the Responsibility scale differentiates individuals who are unable to stick to a task.

Received November 15, 1957.

Reference

Gordon, L. V. *Gordon Personal Profile.* Yonkers, N. Y.: World Book, 1953.

---

[1] The views expressed here are not to be construed as necessarily reflecting those of the Navy.

Journal of Consulting Psychology
Vol. 22, No. 5, 1958

The Effects of Chlorpromazine on Learning in Chronic Psychotics[1]

William A. Whitehead[2] and Leland E. Thune
Vanderbilt University

This study represents an attempt to explore the effects of the drug chlorpromazine on several different aspects of the learning process and on verbalized social adaptation as these take place in a group of hospitalized chronic psychotic subjects (*S*s). The learning variables sampled included serial verbal learning, retention of serial verbal material learned on the previous day, the acquisition of a motor skill, reminiscence of this motor skill, problem solving, and a somewhat related task involving verbalized social adaptation and judgment. These tasks would seem to tap various levels of the learning process, ranging from simple motor learning to problem solving and, finally, "social adaptation." Inasmuch as this was an exploratory study, no directional hypotheses were offered.

Method

Subjects and Treatment

Fifty-two chronic psychotic male veterans, committed to a Veterans Administration Hospital, were studied. *S*s were chosen by the medical staff on the basis of judgments that they were physically suitable for chlorpromazine treatment, that they would be able to cooperate when tested with a battery of simple psychological tests, and that they were free from organic disturbances which might influence their behavior. The majority of the *S*s carried the diagnosis of one of the subtypes of schizophrenia, but there were several who were classified as manic-depressive. *S*s were on the Continued Treatment Service where they received the same routine ward care to which they had become accustomed. None of the *S*s received any other type of somatic therapy during the period of study, but several *S*s were members of an activity group. Prior to treatment, *S*s were randomly divided into two groups; one of which was later designated as an experimental group and the other as a control group. The physician who took charge of ordering the medication arbitrarily designated these groups. No other hospital personnel were aware of which was the experimental group and which was the control group. The mean age for the experimental group was 38 and for the control group was 39. The mean duration of hospitalization was 8.9 years for the experimental group and 7.8 years for the control group.

All *S*s in the experimental group were placed on oral chlorpromazine for a period of two months. Dosage was gradually increased to about 800 mg per day, usually over a period of about 35 days, and then gradually decreased until evaluation, at which time 36 *S*s were receiving a minimum of 300 mg, and 16 *S*s were receiving 200 mg per day. Equal numbers of control and experimental *S*s were represented in each minimum dosage group. *S*s in the control group received placebos prepared by the Smith, Kline, and French Drug Co. All *S*s received the same routine laboratory procedures.

---

[1] This article is based upon part of a dissertation submitted to the Graduate School, Vanderbilt University, in partial fulfillment of the requirements for the Ph.D. degree by the senior author and written under the direction of the junior author. The study was carried out at the Veterans Administration Hospital, Murfreesboro, Tennessee. Much aid was given by F. H. Deter, H. B. McTyre, and other members of the hospital staff.

[2] Now at Central State Hospital, Nashville, Tennessee.

Procedure

All *Ss* were tested individually in two sessions on two consecutive days.

A pursuit rotor apparatus was used to measure motor skill acquisition. The turntable was set to run in a clockwise direction at 40 rpm. A chronoscope measured the "time on target." Instructions were followed by 8 trials of 45 sec. each, with a 15-sec. rest interval between trials. If *S* refused to do the task in spite of urging, a subjective judgment was made as to whether he would be able to cooperate on any of the other tasks. Sometimes the testing procedure had to be stopped at this point.

A Lane-type memory drum was used in the verbal learning and retention tasks. The following 8-item list was learned by the serial anticipation method: *sand, tool, dust, inch, pain, robe, junk, year*. Each word was exposed for 3 sec., and individual trials were separated by a 6-sec. rest interval. The word list was exposed 20 times, or until it had been learned to a criterion of three consecutive complete repetitions, whichever occurred first.

After completing the serial learning task, *S* returned to the pursuit rotor where he received five additional trials.

A questionnaire called the Verbalized Social Adaptation Questionnaire was developed specifically for the study.[3] The questionnaire consisted of a series of statements which presumably contained a principle by which social cooperation could be elicited. Each statement was followed by a question with two alternative answers. The answer which made use of the "principle" given in the statement was considered correct. Following is an example of one of the items: "A person who is lost in a strange city can often find out how to get where he wants to go by asking someone who lives there. If you were lost in a strange town, would you: (*a*) Ask directions from a policeman, or (*b*) Look carefully at all the street signs until you found the one you wanted?" Thirty-six such items were formulated. The 36 questions were divided into two forms of 18 questions each. The items were read to each *S* individually. It was felt that a "correct" answer implied a higher level of social adaptation, or judgment, than did an "incorrect" answer. One-half of the *Ss* received Form I of the questionnaire on the first day, and the other half received Form II on the first day.

Upon his return to the testing room on the second day, *S* was first given 10 additional serial anticipation trials, hereafter called retention trials, with the same word list which he had learned the previous day.

After completing this task, *S* worked on the Making the Last Draw Problem (3, p. 140). This problem was chosen because of its comparative simplicity, and its difficulty level is easily altered. The first phase of the problem used four matchsticks as material. If the *S* was able to solve this task three consecutive times out of five trials, he was asked whether or not he could give a rule whereby he could tell someone else how to win the game every time. Those who were unable to solve the problem three consecutive times were told the principle and then given three additional trials. For *Ss* who were able to solve the first problem, the next level utilized five matchsticks as materials; those who failed the first problem, even with help, did not continue to the second level. The five-stick problem was administered in the same manner as the four-stick problem. For those who solved the second problem, the last problem utilized seven matchsticks, and it was administered in the same way the first two problems had been.

The last measurement taken on the second day was the alternate form of the social adaptation questionnaire, and it was administered in the same manner as it had been the day before.

Results

During the course of the testing it was immediately apparent that many *Ss* could not perform on one or more of the tasks. Some *Ss* were able to cooperate on one of the tasks and not on another. The judgment as to whether or not an *S* was testable was made on a subjective basis. The judgment was, however, based on the observed behavior of the *S*, and if there was any question as to his fitness for undertaking a given task, an attempt was made to gain his cooperation. At the time of the testing, *E* was not aware as to which individuals were being given the drug and which were receiving placebos.

It is felt that the "testable–not testable" factor is itself an important source of information as to the broader behavioral effects of the experimental drug. A more direct measure of the effect of chlorpromazine on learning performance per se is obtained by comparing the performance records for those *Ss* in each group who were able to undertake each task.

Serial Verbal Learning

Twenty *Ss* in the chlorpromazine group were testable on the verbal learning task, while only nine of the placebo group *Ss* were testable. The resulting chi-square figure of 9.6 is significant at less than the .01 level.

When testable *Ss* in each group were compared on individual original learning and retention learning trials, none of the results approached significance. There was also no significant difference between groups when the sums of individual scores were compared. The Median Test was used to make these comparisons.

---

[3] George E. Copple initiated the concept of the Verbalized Social Adaptation Questionnaire and developed several of the items. A copy of the questionnaire has been deposited with the American Documentation Institute. Order Document No. 5583, remitting $1.25 for 35-mm. microfilm or $1.25 for 6 by 8 in. photocopies.

Motor Skill Acquisition

A significantly larger number of *S*s in the chlorpromazine group, 24, than in the placebo group, 16, were found to be able to undertake the pursuit rotor task ($\chi^2 = 6.9$, $p < .01$).

When the nonparametric White's Test was applied to each of the trials and also to the sums of individual scores on all of the trials for both the original learning phase and the reminiscence phase, the results did not indicate a significant difference between groups.

Problem Solving

Twenty-one chlorpromazine *S*s were testable on the problem-solving task, while only 15 placebo *S*s were testable. However, the resulting chi square of 3.6 is not considered significant for purposes of this study. No significant differences were obtained when the groups were compared on the number of correct solutions to each of the problems by means of the chi-square test.

Verbalized Social Adaptation

When the two groups were compared on the testable–not testable split by means of the chi-square test, significantly more *S*s in the chlorpromazine group, 23, were testable than in the placebo group, 16 ($\chi^2 = 6.3$, $p < .02$).

In analyzing the data in this section, the Median Test was used in preference to White's Test because the large number of ties precluded the efficient application of White's Test. When the groups were compared with regard to the number of correct items on both forms of the questionnaire, the groups differed at less than the .05 level. The groups were also significantly different when compared on each form separately. In all instances, the chlorpromazine group had significantly more items correct.

Table 1

Summary of Statistical Findings

Task	Testable—not testable	Testable *S*'s only[a]
Serial verbal learning		
Original	$\chi^2 < .01$	MT NS
Retention	$\chi^2 < .01$	MT NS
Pursuit rotor learning		
Original	$\chi^2 < .01$	WT NS
Reminiscence	$\chi^2 < .01$	WT NS
Problem solving		
Four sticks	χ^2 NS	χ^2 NS
Five sticks	χ^2 NS	χ^2 NS
Seven sticks	χ^2 NS	χ^2 NS
Verbalized Social Adaptation		
Form I		MT<.05
Form II		MT<.05
Combined forms	$\chi^2 < .02$	MT<.05

[a] The test employed and the obtained level of significance is indicated in each case. MT indicates Median Test; WT indicates White's Test; NS indicates not significant at the .05 level.

In order to obtain some check on the reliability of the two forms of the questionnaire and also to get some indication of the reliability of the individual's responses to the two forms, a Pearson correlation coefficient between the two forms was calculated. The resulting coefficient of .85 is significant at less than the .01 level.

Summary of Statistical Findings

Table 1 presents a brief summary of the statistical findings. Rather than presenting the findings for the individual trials on the verbal and motor learning tasks, the table shows the results for the sums of individual scores on all trials combined. The other table headings refer to previously mentioned tests.

Discussion

Effects of Chlorpromazine on Level of Cooperation

Because of the differences between the two groups on the testable–not testable factor on all of the tasks except the problem-solving task, it is inferred that chlorpromazine increases the motivational level of chronic psychotic patients to the extent that they are willing (able) to try the tasks which were presented. Although the drug seems to have little effect on the learning processes themselves, it appears to have a definite effect on motivational factors associated with the process of learning.

One of the reasons that previous studies have indicated such a large discrepancy between the learning of psychotic groups and normal control groups may be that most research workers have not adequately controlled for differences in levels of motivation and cooperation. The findings of the present study regarding the differences between the two groups on the testability factor would seem to lend some support to the hypothesis that motivation is one of the key factors in the learning deficit found in psychotic groups. It may not be that the chronic psychotic patients are unable to learn as efficiently as normals but, rather, that they are more handicapped by the pressure to reduce anxiety or, as Hunt (2) suggested, social approval has lost its reward value.

Effect of Chlorpromazine on Learning Tasks

There were no significant differences between the two groups on the serial verbal learning task. This finding suggests that chlorpromazine does not significantly affect the learning processes involved in verbal learning but, rather, that it does have some influence on the motivational aspects of such learning. It was further noted that the learning curves for these samples of chronic psychotic *S*s were similar to those reported with other types of *S*s.

The results of the present study would seem to indicate that chlorpromazine has little effect on the acquisition of simple motor skills among those *S*s who were able to undertake the task. In spite of clinical reports to the effect that chlorpromazine induces general psychomotor retardation, this retardation did not significantly decrease the motor learning ability of this group. It may be that there was a two-factor effect operating on the motor task such that the psychomotor retardation produced by chlorpromazine lowered the rate of learning while the increased motivation and interest enabled the group to do as well as the control group. It was also noted during the course of the testing that several *S*s whose behavior and appearance was "Parkinsonian" were still able to learn the motor task at about the same rate as other *S*s.

On the problem-solving task, the groups did not differ significantly even on the testable–not testable split. It will be noted, however, that the groups differ at the .10 level in the direction that more *S*s in the chlorpromazine group were testable. Apparently, chlorpromazine has little effect on this particular problem-solving task.

Verbalized Social Adaptation

There were statistically significant differences between the two groups on all phases of the Verbalized Social Adaptation Questionnaire. It should be recognized that the only validity claimed for this instrument is its face validity.

In formulating the rationale for the questionnaire it was felt that the items might have some relevance to what might be called "readiness for psychotherapy." The ability to make generalizations from the therapeutic situation to a particular problem in life would seem to be a necessary, but not a sufficient, requisite for therapeutic progress. It would seem that patients who were unable to correctly answer most of the items on this questionnaire would be poorer therapeutic risks than would those patients who answered most of the items correctly. If one accepts this rationale, then the conclusion that psychotic patients receiving chlorpromazine are more accessible to psychotherapy does not seem unwarranted.

A secondary rationale for the relevance of the items to "readiness for psychotherapy" is that those individuals who are more aware of and concerned with their social environment would seem to be able to form a more satisfactory therapeutic relationship than patients who do not have this interest. One of the most difficult problems in doing psychotherapy with chronic psychotic patients is to arouse their motivation and interest in the social environment to the extent that it will be possible to build a therapeutic relationship. The findings on this questionnaire would seem to indicate that chronic psychotic patients receiving chlorpromazine are more concerned with their social environment.

Implications for Psychotherapy

If it is assumed that chlorpromazine exercises its beneficial effects by reducing overall anxiety, then some speculations regarding theoretical implications for psychotherapy may be made. If the psychotherapeutic interview is viewed as a learning situation, and if the learning tasks used in this study have any relation to the type of learning that takes place in therapy, these data should offer some information which may be integrated into the theory of anxiety reduction in psychotherapy.

Assuming that chlorpromazine reduces anxiety in the chronic psychotic, the finding that there were no significant differences between the two groups on the verbal, motor, and problem-solving tasks seems to be contrary to the hypothesis that a reduction in anxiety level would result in more effective learning. However, the finding that the chlorpromazine groups are more cooperative and seemingly more interested in their social environment suggests that while anxiety reduction may not directly influence the learning processes, it may influence other relevant factors which make new learning possible.

Summary

The effects of chlorpromazine upon the performance of chronic psychotic *S*s on several learning and retention tasks were studied. Fifty-two *S*s were randomly divided into two groups, an experimental group which received chlorpromazine for a period of two months and a control group which received placebos. At the end of this period *S*s were examined on a serial verbal learning task, a motor learning task, a problem-solving task, and on a questionnaire called the Verbalized Social Adaptation Questionnaire. The following conclusions were drawn:

1. Chlorpromazine appears to have little effect on the learning processes per se.
2. Chlorpromazine has a significant effect on improving the responses of chronic psychotics to items judged to involve social adaptation.
3. Chlorpromazine significantly increases the number of chronic psychotic *S*s who are motivated to cooperate in the testing procedure.
4. In spite of reports to the effect that chlorpromazine produces general psychomotor retardation, *S*s receiving the drug learn a motor task at the same rate as *S*s who are not receiving the drug.

Received August 15, 1957.

References

1. Edwards, A. *Statistical methods for the behavioral sciences.* New York: Rinehart, 1955.
2. Hunt, J. McV., & Cofer, C. N. Psychological deficit. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald Press, 1944. Pp. 971–1032.
3. Ray, W. S. Complex tasks for use in human problem-solving research. *Psychol. Bull.,* 1955, **52**, 134–149.

Journal of Consulting Psychology
Vol. 22, No. 5, 1958

Stylus Maze Performance of Chronic Schizophrenics Taking Chlorpromazine[1]

Paul G. Daston[2]

Veterans Administration Hospital, Brockton, Massachusetts

Conflicting reports have been published regarding Porteus maze performance of chronic schizophrenics treated with chlorpromazine (CPZ). The present study employs a stylus maze to investigate effects large doses of CPZ may have on performance of chronic schizophrenics when practice effect is minimized.

*S*s were 16 chronic schizophrenics, divided into CPZ (1200 mg. daily) and placebo (PL) groups. Ritalin[3] was added for CPZ *S*s the first 15 days and for PL *S*s instead the last 4 days. All had 27 previous weekly trials, with performances approaching an asymptote. Groups were equated for performance and tested daily for 19 days. Apparatus was a Lafayette stylus maze, wired to record errors. The stylus traced the path taken on a paper schematic placed under the maze, thus recording extent of blind alley penetration. Time was kept by stop watch. *S*s were told to take the shortest path to the goal without touching the sides, to avoid blind alleys and to work as fast as possible.

As time and error scores were related (*S*s taking more time tended to make fewer errors), each measure was transformed into standard scores and both combined. Analyses of variance, using difference scores, revealed no significant differences between groups, either for 1st vs. 15th day, 15th vs. 19th, or 1st vs. 19th day. In no case did the F ratio exceed 1, indicating CPZ, with or without Ritalin, had little effect on performance or, inferentially, perceptual-motor coordination. However, inspection of means revealed CPZ group performance remained about the same with or without Ritalin, whereas PL mean scores improved with its addition. Also, chi-square tests of direction of change showed a significant shift ($p = .05$) favoring PL improvement.

It was felt extent of blind alley penetration might give a measure of judgment. A number score was assigned to each 16th inch of blind alley penetration. Statistically significant differences were found between groups for the first measure, so analysis of covariance was used. Differences were not significant for either 15th or 19th day, indicating judgment was not affected differentially by CPZ.

In summary, perceptual-motor coordination, as measured, was not affected significantly by large doses of CPZ. PL errors tended to decrease. Judgment, denoted by extent of penetration into blind alleys, was not differentially affected. Ritalin effect warrants further study.

Brief Report.
Received May 19, 1958.

---

[1] An extended copy of this report may be obtained without charge from Paul G. Daston, Clinical Psychology Section, Veterans Administration Hospital, Durham, N. C., or for a fee from the American Documentation Institute. Order Document No. 5691, remitting $1.25 for microfilm or $1.25 for photocopies.

[2] Now Chief Psychologist at Durham (N. C.) Veterans Administration Hospital.

[3] Ritalin, a "psychic energizer," purports to minimize such side effects of CPZ as drowsiness and mood depression.

Journal of Consulting Psychology
Vol. 22, No. 5, 1958

The Effect of Differential Motivating Instructions on the Emotional Tone and Outcome of TAT Stories[1]

Harriet C. Sumerwell, Mary M. Campbell,
and Irwin G. Sarason[2]
University of Washington

This study was carried out to achieve two goals. The first was to determine the extent to which we could replicate certain of Eron's (1, 2, 3) findings concerning his scales for scoring Emotional Tone and Outcome on the TAT. In particular, we were interested in whether or not the reliabilities of raters and pull values of the TAT cards in our study would approximate those obtained by Eron in his normative study.

Our second objective was to assess the sensitivity of Eron's Emotional Tone (ET) and Outcome (O) scales to experimental instructions introduced prior to obtaining subjects' stories to the cards. One of the widely used sets of TAT instructions in research and clinical work is that of Murray (4). Therefore Murray's instructions (adapted for group administration) constituted one of four sets of instructions used in this experiment. In examining Murray's instructions, it seemed to us that they either directly or indirectly provide the subject with information which might lead him to draw the inference that the test he is about to take is some combination of an intelligence and personality test. We attempted to separate these two possible effects by employing another set of instructions informing the subject he was about to take an intelligence test and still another set informing the subject he was to take a personality test. Our fourth set consisted of neutral instructions which merely informed the subject of the procedure he was to follow in responding to the cards.

For the present purposes we assumed that the ET and O ratings would reflect the degree of stress experienced by the subject and that the lower the ratings (i.e., greater dysphoria) the greater was the subject's reaction to stress. Our hypothesis concerning the effects of the differential instructions was that the Murray instructions would be the most stressful of the four sets used and that the neutral instructions would be the least stressful. Consequently, we expected relatively low ET and O ratings under the Murray instructions and relatively high ratings under the neutral instructions. The ET and O ratings for the Projective and Intelligence instructions were expected to fall between those obtained under the Murray and Neutral instructions.

Method

Subjects

The *S*s were 106 volunteers from introductory psychology classes at the University of Washington. The total number of *S*s in each group was as follows: Murray—20, Intelligence—22, Projective—42, and Neutral—22. Half the *S*s in each group were tested by one *E*, and half by the other *E*. There were equal numbers of males and females in the Murray, Intelligence, and Neutral groups, although we had made no systematic attempt to control for the sex variable. In the Projective group, females outnumbered males 29–13.

Procedure

Eight TAT cards—1, 2, 4, 6BM, 13MF, 16, 18GF, and 18BM—were presented to each group of *S*s by projection upon a screen after the appropriate instructions had been given. Each picture was exposed for five minutes, during which time the *S*s wrote their stories. The four sets of instructions were administered as follows:

---

[1] Portions of this paper were presented at the 1957 meeting of the Western Psychological Association.

[2] Irwin G. Sarason is responsible for the manuscript in its present form.

1. *Murray instructions.* This is a test of imagination, one form of intelligence. I am going to show you some pictures, one at a time; and your task will be to make up as dramatic a story as you can for each. Tell what has led up to the event shown in the picture, describe what is happening at the moment, what the characters are feeling and thinking; and then give the outcome. Write your thoughts as they come to your mind. Do you understand? Here is the first picture.
2. *Intelligence test instructions.* Being college students you are probably accustomed to taking intelligence tests. This is a special test of intelligence and the ability to think creatively. Rather than measuring specific abilities, this test gets at innate capacity and potentialities. I am going to show you certain pictures, one at time; and your task will be to make up a story for each. Tell what has led up to the event shown in the picture, describe what is happening at the moment, what the characters are feeling and thinking; and then give the outcome. Write your thoughts as they come to your mind. Do you understand? Here is the first picture.
3. *Personality test instructions.* You have probably heard of projective techniques and tests. The test you are about to take is the Thematic Apperception Test. It is a projective test in the sense that you unconsciously project your inner feelings and motives into the stories you tell. I am going to show you some pictures, one at a time; and your task will be to make up a story for each. Tell what has led up to the event shown in the picture, describe what is happening at the moment, what the characters are feeling and thinking; and then give the outcome. Write your thoughts as they come to your mind. Do you understand? Here is the first picture.
4. *Neutral instructions.* I am going to show you some pictures, one at a time; and your task will be to make up a story for each. Tell what has led up to the event shown in the picture, describe what is happening at the moment, what the characters are feeling and thinking; and then give the outcome. Write your thoughts as they come to your mind. Do you understand? Here is the first picture.

The four sets of instructions were group-administered by two different *E*s. Thus there were eight experimental groups in the experiment, and it became possible to evaluate the effects of the experimenter difference.

After all the data had been gathered, the stories for all *S*s were independently scored by the two *E*s for ET and O. The scoring procedure was so arranged that neither *E* had knowledge of the group to which the *S*, whose protocol was being scored, belonged. Disagreement between raters was resolved for the purposes of the statistical analysis of the data by the third author (Sarason) who had had considerable experience in using Eron's scales.

Results

Pearson product-moment correlations between the ratings (made prior to resolution of disagreement) of the two examiners were .817 for ET and .808 for O. These correlations are similar to those obtained by Eron (1, 2, 3) and seem to justify increased confidence in the use of this scoring system for experimental purposes.[3] In general, all of the cards tended to elicit stories negative in both ET and O, a finding which again corroborated Eron's and Sarason and Sarason's earlier results.

The effects of the experimental instructions and the two examiners were evaluated by means of analysis of variance. The measure employed for each *S* was the sum of the ratings over all cards administered. Table 1 presents the means and *SD*s of ET and O ratings as a function of the experimental conditions. Preliminary comparisons of the eight experimental groups indicated that the assumption of homogeneity of variance could be met in our statistical tests.

In general, no significant difference between examiners was found for either the ET or O ratings. There was, however, a significant difference at the .01 level between examiners for O under Neutral instructions.

Since there were no significant differences between examiners for ET, the data were pooled into the four instruction groups. A significant difference at the .01 level among these groups was found. Significant differences between the four pooled groups under O were also obtained at the .01 level. However, since significant examiner differences under the Neutral condition for the O ratings had been

---

[3] The correlations obtained are also similar to those obtained by Sarason, I. G. and Sarason, B. R. in an unpublished study, "The effect of type of administration and sex of subject on emotional tone and outcome ratings of TAT stories."

Table 1

Means and Standard Deviations of Emotional Tone and Outcome Ratings as a Function of Instructions and Examiners

Measure and instructions	Examiner No. 1		Examiner No. 2		Average of 2 Examiners	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Emotional Tone						
Murray	−8.90	2.02	−9.20	2.39	−9.05	2.16
Projective	−8.57	2.36	−8.29	2.69	−8.43	2.50
Intelligence	−7.73	2.41	−8.55	2.34	−8.14	2.36
Neutral	−6.00	1.84	−6.73	2.28	−6.36	2.06
Outcome						
Murray	−4.10	3.70	−4.80	3.85	−4.45	3.69
Projective	−3.24	3.27	−2.19	4.04	−2.71	3.67
Intelligence	−3.45	3.83	−1.73	3.74	−2.59	3.80
Neutral	+ .82	3.25	−1.91	2.02	− .55	2.99

Note.—The measure for each *S* was the sum of ratings over all cards administered.

found, a further analysis of variance of O ratings was performed omitting the Neutral group. When the three remaining groups were compared, no significant differences were discovered among the three remaining groups.

When *t* tests were performed, it was found that for the ET ratings, the Murray, Projective, and Intelligence groups differed significantly from the Neutral group at the .01, .01, and .05 levels of confidence respectively. The Murray, Projective, and Intelligence groups did not differ significantly among themselves.

With respect to the O rating, the Murray and Projective groups were found to differ significantly from the Neutral group (.01 and .05 levels respectively). The Intelligence group did not differ significantly from the Neutral group. As already mentioned, the Murray, Projective, and Intelligence groups did not differ among themselves on the outcome ratings.

It seemed of interest to compute the correlation between ET and O. This *r* was found to be .385, which was low but significant at the .01 level. The small size of this correlation seems to be accounted for by the greater variability of O ratings, similarly reported by Eron, and the observation that *S*s occasionally tended to give a happy ending quite out of line with the initial depressing mood of their story.

A separate analysis of card differences was also performed. It was found that the scores for individual cards differed significantly beyond the .001 point for ET under the four instructional conditions. For O, there were significant card differences at the .001 point for Murray, Personality, and Intelligence groups, but no significant differences under the Neutral group were found. This again appeared to be related to examiner differences under the Neutral condition.

In testing for the effect of instructions upon cards, the coefficient of concordance was used, the cards being ranked on the basis of their mean scores obtained under each of the experimental conditions. It was found that instructions did not alter the ranks of the cards significantly. Thus it appears that a card retains its original pull value relative to the other cards in spite of the over-all effect of varied instructions.

Discussion

With respect to the hypotheses we made concerning our experimental procedure, it appears as though they were supported by our results. Happier emotional tone and outcome were more associated with the Neutral instructions than with the other instructions employed. Although this study does not provide sufficient basis for delineating precisely the reactions of the *S*s to the instructions, it seems reasonable that negative ET and O scores are associated with emotional upset, tension, and dysphoria. In this regard we wonder whether instructions such as Murray's do not impose undue stress and structure on the testing situation. We generally interpret *S*'s reactions to the TAT in terms of the test stimuli presented to him and in terms of characteristics peculiar to the individual. However, there may be a third variable we overlook in test administration, namely, the influence of our preliminary instructions on *S*'s behavior. If our aim in testing is to reduce the degree of structure in the test situation, perhaps instructions similar to those we would call Neutral would be more appropriate.

Our results have provided further basis for employing Eron's rating scales for research purposes. With respect to reliability and other characteristics, our results are generally consistent with those reported by others. It would seem useful in future research to study the relationship of these and other rating scales to a variety of behavioral criteria and experimental procedures.

Frequently widely different stories achieved similar ratings on Eron's scale. We feel that other rating scales designed to tap a variety of personality dimensions and content areas might yield pertinent information about *S*s' differences to which the ET and O scales are relatively insensitive. Further research might be directed toward devising other scales and later towards investigating the meaning of *S* differences along these dimensions.

Summary

The effects of four kinds of instructions and of two different examiners on TAT emotional tone and outcome ratings were studied. One group of *S*s received Murray's instructions; one a set of instructions informing *S*s they were about to take a test of intelligence; another a set of instructions informing *S*s they were about to take a personality test; and a fourth group of *S*s received neutral preliminary instructions.

Adequate interrater reliability for emotional tone and outcome was obtained. A difference due to examiners was found only for outcome ratings under the neutral condition.

The Murray, personality, and intelligence instructions led to more depressive, sadder stories than did the neutral instructions. The implications of this finding with respect to TAT testings were discussed. It was concluded that neutral instructions might be the most appropriate kind of instructions to give *S*s prior to taking the TAT.

Received September 23, 1957.

References

1. Eron, L. D. A normative study of the Thematic Apperception Test. *Psychol. Monogr.*, 1950, 64, No. 9 (Whole No. 315).
2. Eron, L. D., & Ritter, Anne M. The comparison of two methods of administration of the Thematic Apperception Test. *J. consult. Psychol.*, 1951, **15**, 55-61.
3. Eron, L. D., Terry, Dorothy, & Callahan, R. The use of rating scales for emotional tone of TAT stories. *J. consult. Psychol.*, 1950, **14**, 473-478.
4. Murray, H. A. *Thematic Apperception Test.* Cambridge, Mass.: Harvard Univer. Press, 1943.

Journal of Consulting Psychology
Vol. 22, No. 5, 1958

Deviant Responses as Indicators of Immaturity and Schizophrenia

S. O. Hesterly and Irwin A. Berg
Louisiana State University

When patterns of responses to various stimuli are examined, the responses often are biased and do not follow a normal probability distribution. Such "set" or *einstellung* phenomena tend to appear most clearly as the stimulus pattern is unstructured. There are myriad forms of such biased responses, such as the tendency to call "heads" on the first toss of a coin, the tendency to turn right when one may turn left or right, etc. (8, 14). In psychological tests, Cronbach (12, 13) has identified a large number of such biases and called them *response sets.*

Since there is evidence that such sets or biases are expressions of personality traits (1, 3, 6, 11, 15, 19, 21, 22, 23, 24), Berg (5) sought to employ these patterns of bias as measures of personality but with only limited success. However, when attention was shifted from the pattern of bias to those responses which departed from the established set or bias, it was found possible to use these departures as measures of personality and other characteristics. Barnes (2), for example, was able to construct a series of clinical scales using this concept of deviant responses. Thus the key to the problem of utilizing response bias resides in using, not the pattern of bias itself but, rather, the responses which go counter to the established bias, i.e., the deviant responses. In other words, we should not pay particular attention to the 80% of the population who call "heads" when a coin is flipped, but we would closely scrutinize the 20% who call "tails." The latter response, in our lexicon, is deviant. A fuller account of the problem and its background is given in articles by Berg (4, 5).

The present study is one of a series of empirical tests of the Deviation Hypothesis which has been stated as follows: "Deviant response patterns tend to be general; hence those deviant behavior patterns which are significant for abnormality (atypicalness) and thus regarded as symptoms (earmarks or signs) are associated with other deviant response patterns which are in noncritical areas of behavior and which are not regarded as symptoms of personality aberration (nor as symptoms, signs, earmarks)" (5, p. 159). In other words, by using deviant responses in a noncritical area, such as a liking for designs, we should be able to measure deviant behavior in a critical area such as psychopathology, chronic organic disease, employee morale, creativeness, etc.

The specific hypotheses of the present study were two:

1. The response patterns of normal, young children to the Perceptual Reaction Test (PRT) (7) are significantly different from those of normal adults. As a corollary, older children have response patterns more similar to those of adults than younger children.

2. The response patterns of normal, young children on the PRT more nearly approximate those of adult schizophrenics than those of normal adults.

Thus maturity as a critical area of behavior is assessed by means of deviant response patterns as they occur in the PRT. Then, since descriptions of schizophrenic behavior (9, 10, 16, 17, 18, 20) typically make reference to immaturity as a common feature of the schizophrenic reaction, it is predicted that the deviant response patterns of normal, young children are similar to those found in schizophrenics.

Table 1

Distribution of Subjects by Age, Sex, and Place of Residence

Age and residence	7–0 to 8–11	9–0 to 10–11	11–0 to 12–11	Totals
Male				
Urban	25	25	25	75
Rural	25	25	25	75
Female				
Urban	25	25	25	75
Rural	25	25	25	75
Totals	100	100	100	300

Procedure

In order to elicit deviant response patterns, the PRT was used. This test consists of 60 abstract designs drawn with ruler and compass. The subject is required to mark one of the following options for each design: *like much* (LM), *like slightly* (LS), *dislike slightly* (DS), or *dislike much* (DM). No obvious meaning is inherent in any of the designs; hence, the test is considered to be relatively unstructured, a condition which facilitates the appearance of response biases. It should be noted that other stimulus patterns could be employed. However, the PRT has been used as a research instrument for a number of years and, accordingly, a large body of normative data is available. Also, Barnes' (2) data on PRT responses of adult schizophrenics could be used.

Three hundred Louisiana grade-school children were used as *S*s for the present study. See Table 1. The PRT was administered to the *S*s in their regular classrooms during a normal school day. Administration time was approximately 12 minutes. The *S*s' responses were transferred from the PRT booklets to IBM answer sheets for machine scoring. The schizophrenia scales developed by Barnes (called *Sigma Scales* by him) were used to determine each individual's score. Barnes' scales were developed by assigning plus weights to those PRT responses which were significantly characteristic of his schizophrenic group and assigning minus weights to PRT responses which were characteristic of his normal group. The sum of these weights was the schizophrenic score. These scores were compared on the basis of age and place of residence by analysis of variance. The resulting variances for the different groups appeared to be rather heterogeneous; therefore, Bartlett's test for homogeneity of variance was applied to the data. Since all of the resulting chi-square values were statistically significant beyond the .01 level of probability, a nonparametric technique was used to compare the group schizophrenia scores. Specifically, the nonparametric median test seemed to apply to data of the type obtained, as it is not dependent upon homogeneity of variance. The hypothesis tested is that the groups compared are random samples from a population with a common median.

Results

Median schizophrenia scores on the PRT are given for all *S*s in Table 2. The results of the nonparametric median tests for male *S*s are shown in Table 3. The first comparisons are between normal, male children and normal, adult males. The resulting chi-square value of 138.14, which is statistically significant at the .0001 level of probability, supports the hypothesis that normal male children attain schizophrenia scores on the PRT that differ significantly from those of normal, adult males. The second comparison was between normal, male children and schizophrenic adult males. The obtained chi-square value of

Table 2

Median Schizophrenia Scores on the PRT for 300 Normal Children, 850 Normal Adults, and 167 Schizophrenic Adults Taken Separately and as Combined for Nonparametric Median Tests

	Median Score		
Subjects	Group only	Combined with normal adults	Combined with schizophrenic adults
Normal children			
Males ($N=150$)	+ 8.50	−14.15	+5.80
Females ($N=150$)	+ 3.25	− 9.62	+3.36
Normal adults[a]			
Males ($N=500$)	−17.83		
Females ($N=350$)	−12.81		
Schizophrenic adults[a]			
Males ($N=99$)	+ 1.40		
Females ($N=68$)	+ 3.50		

[a] The data for normal adults and schizophrenic adults are from Barnes (2).

Table 3

Nonparametric Median Test for PRT Schizophrenia Score Differences Between Groups of Male Subjects

	Number						
Subjects	Below median	Above median	Total	Subgroup median	Chi square	*df*	*p*
Normal children	10	140	150	+ 8.50	138.14	1	<.0001
Normal adults	309	191	500	−17.83			
Normal children	65	85	150	+ 8.50	3.10	1	>.05
Adult schizophrenics	55	44	99	+ 1.40			
Normal children							
7–10 to 8–11 years	17	33	50	+16.50	7.10	2	<.05
9–0 to 10–11 years	27	23	50	+ 5.50			
11–0 to 12–11 years	31	19	50	+ 2.50			
Normal children							
Urban	33	42	75	+12.00	1.71	1	>.05
Rural	42	33	75	+ 3.25			

3.10 is not statistically significant, supporting the null hypothesis that the two groups are probably samples from a population with a common median. Next, a comparison was made between three age groups (7–0 to 8–11 years, 9–0 to 10–11 years, 11–0 to 12–11 years) of normal children to determine the possibility of a significant trend in PRT schizophrenia scores as a function of chronological age. The resultant chi-square value of 7.10, which is statistically significant at the .05 level of probability, supports the hypothesis that the schizophrenia scores of normal children decrease significantly with age. The final comparison was made between groups of normal, urban and rural, male children. The

Table 4

Nonparametric Median Test for PRT Schizophrenia Score Differences Between Groups of Female Subjects

	Number						
Subjects	Below median	Above median	Total	Subgroup median	Chi square	*df*	*p*
Normal children	29	121	150	+ 3.25	80.97	1	<.0001
Normal adults	223	127	350	−12.81			
Normal children	76	74	150	+ 3.25	.06	1	>.05
Adult schizophrenics	34	34	68	+ 3.50			
Normal children							
7–0 to 8–11 years	18	32	50	+11.50	6.78	2	<.05
9–0 to 10–11 years	26	24	50	+ 3.00			
11–0 to 12–11 years	32	18	50	− 2.50			
Normal children							
Urban	33	42	75	+ 6.75	2.16	1	>.05
Rural	43	32	75	+ 1.00			

obtained chi-square value of 1.71, which is not statistically significant, supports the null hypothesis that the two groups appear to be samples drawn from a population with a common median.

Table 4 presents the results of the nonparametric median tests performed to compare the median differences in scores for the Barnes schizophrenia scale obtained on the PRT between groups of female *S*s. The comparisons made between groups of female *S*s were the same as those for male *S*s. Although the chi-square values varied somewhat, the probability levels obtained were identical with those obtained for male *S*s. Thus, the conclusions reached regarding females were the same as those for males.

Discussion

The results of the present study support the hypothesis, as stated earlier, that the response patterns of normal, young children to the PRT are significantly different from those of normal adults. Their patterns more nearly approximate those of adult schizophrenics than those of normal adults. As normal children mature their response patterns gradually approach the normal adult patterns for the PRT. Thus it appears that these differential response pattern biases may be used to assess maturity, insofar as the evidence of the present study indicates.

Since normal children make responses to the stimulus patterns of the PRT that are not significantly different from those of adult schizophrenics, it may be inferred that some common factor or factors exist between the two groups. As noted earlier, immaturity of behavior has frequently been used as a phrase characterizing schizophrenics. The results of this investigation support this description, at least insofar as similarity of performance on the PRT of normal children and adult schizophrenics is concerned. Further, since the schizophrenia scores of normal children decrease significantly with age, gradually approaching the normal adult level, the schizophrenia scale could be used as a crude measure of maturity. In fact, there is every reason to believe that with further investigation, using wider age ranges and larger age samples, a reasonably precise measure of maturity could be developed.

As urban and rural children were not significantly different with reference to schizophrenia scores on the PRT, it would probably be unnecessary to develop separate norms. It should be noted, however, that the actual differences in median scores for the two groups, although not statistically significant, were in the same direction for both males and females. Urban children scored slightly higher than rural children. This difference might possibly be found to be significant for larger groups of children.

Summary

The present investigation was undertaken to determine whether groups differing in chronological age could be shown to differ in response patterns on the PRT, as predicted by Berg's Deviation Hypothesis. In addition, the investigation included a comparison of the response patterns for groups of normal children and schizophrenic adults, both being characterized by immaturity of behavior, to determine whether they might be similar. Groups of normal children were compared, separately, with groups of normal adults and schizophrenic adults. Significant differences were found between the response patterns of normal children and those of normal adults. Normal children and schizophrenic adults did not differ significantly.

A significant trend in response patterns for normal children as a function of chronological age was found when three consecutive age groups were compared. Urban and rural children were found to have response patterns on the PRT which were not significantly different. Thus, it was concluded that response patterns on the PRT are related to chronological age, at least in the case of normal children, and that, since the response patterns of normal children and schizophrenic adults are not significantly different, some common factor or factors such as immaturity must exist between the two groups. Finally, the deviant response patterns of normal children were sufficiently clear-cut that they could be used to assess maturity.

Received October 3, 1957.

References

1. Barnes, E. H. Response bias and the MMPI. *J. consult. Psychol.*, 1956, **20**, 371–374.
2. Barnes, E. H. The relationship of biased test responses to psychopathology. *J. abnorm. soc. Psychol.*, 1955, **51**, 286–290.
3. Bass, B. M. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 616–623.
4. Berg, I. A. Response bias and personality: The deviation hypothesis. *J. Psychol.*, 1955, **40**, 61–72.
5. Berg, I. A. Deviant responses and deviant people: The formulation of the deviation hypothesis. *J. counsel. Psychol.*, 1957, **4**, 154–161.
6. Berg, I. A., & Collier, J. Personality and group differences in extreme response sets. *Educ. psychol. Measmt*, 1953, **13**, 164–169.
7. Berg, I. A., & Hunt, W. A. *The perceptual reaction test.* Evanston, Ill.: Author, 1948.
8. Berg, I. A., & Rapaport, G. M. Response bias in an unstructured questionnaire. *J. Psychol.*, 1954, **38**, 475–481.
9. Cameron, N. Deterioration and regression in schizophrenic thinking. *J. abnorm. soc. Psychol.*, 1939, **34**, 265–270.
10. Cameron, N. The functional psychoses. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald Press, 1944. Pp. 861–921.
11. Chapman, L. J., & Campbell, D. T. Response set in the F scale. *J. abnorm. soc. Psychol.*, 1957, **54**, 129–132.
12. Cronbach, L. J. Response sets and test validity. *Educ. psychol. Measmt*, 1946, **6**, 475–494.
13. Cronbach, L. J. Further evidence on response sets and test designs. *Educ. psychol. Measmt*, 1950, **10**, 3–31.
14. Goodfellow, L. D. The human element in probability. *J. gen. Psychol.*, 1940, **33**, 201–205.
15. Guilford, J. P. The validation of an "indecision" score for predicting proficiency of foremen. *J. appl. Psychol.*, 1954, **38**, 224–226.
16. Hanfmann, E., & Kasanin, J. Conceptual thinking in schizophrenia. *Nerv. Ment. Dis. Monogr.*, 1942, No. 67.
17. Hoskins, R. G. *The biology of schizophrenia.* New York: Norton, 1946.
18. Kasanin, J. S. *Language and thought in schizophrenia.* Berkeley: Univer. California Press, 1944.
19. Lorge, I. Gen-like: Halo or reality? *Psychol. Bull.*, 1937, **34**, 545–546.
20. Phillips, L. Developmental theory applied to normal psychopathology percepts. *J. Pers.*, 1954, **22**, 464–474.
21. Rapaport, G. M., & Berg, I. A. Response sets in a multiple-choice test. *Educ. psychol. Measmt*, 1955, **15**, 58–62.
22. Rubin-Rabson, G. Correlates of the noncommittal test-item response. *J. clin. Psychol.*, 1954, **10**, 93–95.
23. Voth, A. C. An experimental study of mental patients through the autokinetic phenomenon. *Amer. J. Psychiat.*, 1947, **103**, 793–805.
24. Wallen, R. Food aversions of normal and neurotic males. *J. abnorm. soc. Psychol.*, 1945, **40**, 77–81.

Journal of Consulting Psychology
Vol. 22, No. 5, 1958

Signs of Homosexuality in Human-Figure Drawings[1]

Armin Grams
DePaul University [2]

and Lawrence Rinder
Illinois State Training School for Boys [3]

This study investigates the validity of the 15 signs in human-figure drawing which Machover (1) lists as predictive of homosexuality. Fifty adolescent inmates of a state training school were divided on the basis of homosexual experience into two groups matched for age, schooling, IQ, and race. Each subject was asked first to draw a person and then to draw a person of the sex opposite that of the first drawn figure. Three psychologists rated the drawings for the presence of signs purported to be indicative of homosexuality. To facilitate rating, Machover's signs were stated objectively as follows:

1. Ear large or heavily lined or much detail
2. Detectable delineation of hips or buttocks
3. Failure to complete drawing below waist
4. Heavy line of demarcation at waist
5. Failure to draw "V" of crotch
6. Presence of shading on lips
7. Pants transparent (legs showing through)
8. Naked presence of sexual organs (genitals only)
9. Trousers only clothing shaded
10. Female figure transparent below waist
11. Male nose large, erased, and redrawn
12. Phallic foot (length at least three times width and/or shaded tip)
13. Belt shaded and speared to right of figure
14. Presence of eye lashes
15. Drawing of female figure first

The extent to which all judges rated a sign present or absent in a drawing became the index of rating reliability. Agreement on the drawings of homosexuals was 76.5%, on the drawings of controls, 83.1%.

Three of the signs [1, 2, 4] resulted in chi squares whose P values were between .10 and .20, one [5] resulted in a chi square with a P value between .20 and .30, and three [12, 14, 15] resulted in chi squares whose P values were between .50 and .70. The eight remaining signs resulted in chi squares whose P values exceeded .70. Thus, none of the signs proved to have individual validity. To determine the predictive significance of the 15 signs taken as a group, the total number of signs present in a drawing was correlated with homosexuality and nonhomosexual experience. A point-biserial r of .15, not significant at the .05 level of confidence, was obtained. Thus, neither individually nor collectively did the signs studied validly predict the criterion.

Brief Report.
Received May 1, 1958.

Reference

1. Machover, Karen. *Personality projection in the drawing of the human figure*. Springfield, Ill.: Charles C Thomas, 1949.

---

[1] An extended report of this study may be obtained without charge from Armin Grams, Institute of Child Development and Welfare, Minneapolis 14, Minnesota, or for a fee from the American Documentation Institute. Order Document No. 5691 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies.

[2] Now at Institute of Child Development and Welfare, University of Minnesota.

[3] Now at Blackhawk Mental Health Center, Waterloo, Iowa.

Journal of Consulting Psychology
Vol. 22, No. 5, 1958

An Evaluation of Human-Figure Drawing Indicators of Paranoid Pathology

Marvin Reznikoff and Alma L. Nicholas
Institute of Living, Hartford, Connecticut

The drawing of human figures is a technique which has proved to be of considerable usefulness in the personality evaluation and diagnosis of psychiatric patients. Interpretations of figure drawings, however, are based more often on clinical intuition and experience than on objectively defined or validated criteria. As Holzberg and Wexler (1) point out, in their validation study of figure drawings, this technique remains "much more of an art than a science."

Despite efforts to validate figure drawings, disagreement concerning the significance of certain characteristics of the drawings still exists. Clearly contradictory are some of the findings relating to paranoid indicators. Machover (2) includes, as possible indicators of paranoid traits, eye emphasis, ear emphasis, large figure placed in the middle of the page, a large head, and speared or talon-like fingers. In a recently published study, Ribler (3) investigated two of these indicators, eye and ear emphasis, in paranoid schizophrenics compared with unclassified schizophrenics, anxiety neurotics, and "normals." He found no statistically significant differences between the diagnostic groups with regard to these two variables.

In endeavoring to interpret these discrepant results it is worth noting that Ribler relied solely upon diagnostic categories in classifying his patients as paranoid or nonparanoid. He did not attempt a direct estimation of the paranoid component in a patient's pathology. The intent of the current study was to check some of the drawing features reported as paranoid indicators in a group of patients specifically evaluated for the presence and extent of paranoid symptomatology.

Procedure

Subjects

In selecting the subjects for this study, a number of staff psychiatrists and experienced psychiatric residents at the Institute of Living were asked to indicate any of their presently hospitalized patients who had a history of paranoid behavior and were currently paranoid. They were also asked for a list of patients who had no history of paranoid behavior and currently evidenced no paranoid ideation. When a patient was classified as paranoid, the psychiatrist was requested to rate the degree of paranoid involvement as "mild," "moderate," or "marked." The following definition of paranoid trends, adopted in a modified form from the *Diagnostic and Statistical Manual of Mental Disorders* prepared by the American Psychiatric Association, was used as a guide by the psychiatrists in categorizing their patients:

> Thinking characterized by the use of projection as a defense mechanism. Behavioral manifestations are suspiciousness, envy, extreme jealousy and stubbornness. In more flagrant cases, ideas of reference and delusions of persecution and/or grandeur are evident.

All patients who were under 15 or over 60, of below average intelligence, or having any type of organic involvement were excluded from the study. The actual sample consisted of 61 hospitalized patients, 31 described as paranoid and 30 as nonparanoid. The mean ages of the paranoid and nonparanoid groups were 31.8 and 31.5 years respectively. Both groups contained patients having neurotic and psychotic diagnoses. In the paranoid group, 12 patients were judged to be "mildly paranoid," 9 "moderately paranoid," and 10 "markedly paranoid."

Drawings of both a male and female figure were already available on a number of these patients who had been referred for routine psychological testing prior to this study. The status of these patients at the time of testing was obtained from their physicians. The remaining patients, who had not previously received tests, were individually asked to draw human figures.

Scoring Procedure

A list of 26 carefully defined characteristics of figure drawings, which might discriminate paranoid from nonparanoid patients, was compiled from Machover's indicators, the authors' clinical experience, and relevant items taken from the check list used by Holzberg and Wexler. These items are as follows:

1. Careful detailing of eyelashes
2. Careful detailing of eyebrows
3. Line emphasis on outline of the eyes (reinforcement)
4. Shading of the eyes
5. Two eyes in profile view
6. Eyes represented by circles
7. Eyes represented by crosses
8. Eyes represented by dots
9. Eyes represented by dashes or curves
10. Unseeing eyes; eyes without a pupil
11. Unusual detailing and/or articulation of eye
12. Eyes absent
13. Disproportionately large ears
14. Line emphasis on outline of ears (reinforcement)
15. Ears where none should be present
16. Shading of the ears
17. Ear misplaced in relation to other head features
18. Ears absent
19. Unusual articulation and/or detailing of ear
20. Dark or heavy line emphasis overall
21. Contrasting pressures of line; light and heavy pressures
22. Shading of areas other than eyes and ears
23. Speared or talon-like fingers
24. Disproportionately large head
25. Clothing elaborations to conceal some features of the figure (cloak, cape, any unusual clothing)
26. Size of figure (actual length measurement)

With all 61 sets of drawings randomized and any information relating to group classification removed, the drawings were scored independently by each of the authors. A plus–minus scoring system was used to indicate the presence or absence of the first 25 of the indicators in a patient's set of two drawings. For Item 26, an actual measurement of the first figure drawn was used.

There was 93% agreement between the independent ratings of the two judges. Differences in scoring were reviewed, and agreement was reached with regard to the appropriate score.

Treatment of Data

The number of times each of the first 25 indicators occurred in the two groups was tabulated, and comparisons between the groups were made using chi square. A *t* test was employed to determine whether the paranoid patients differed significantly from the nonparanoids in the size of their figure drawings.

A drawing score was obtained for each patient representing the number of items out of the total of 25 scored present. These scores were converted to percentages and transformed to equalize the variances. A *t* test was then utilized to compare the paranoid and nonparanoid groups. For the paranoid group, the transformed scores were distributed into the "mildly paranoid," "moderately paranoid," and "markedly paranoid" subcategories, and an analysis of variance design was applied to test for differences between these three groups.

Results and Discussion

Of the 25 indicators only 2, line emphasis on outline of eyes (Item 3) and heavy line emphasis overall (Item 20), occurred significantly more frequently in the paranoid group. The chi square value of 6.32 for the eye emphasis item is significant at .02 level. For over-all line emphasis chi square was 4.95, which falls at .05 confidence level. Since 25 tests of significance were performed in this analysis, it should be recognized that even the two significant differences obtained may have been due to chance. According to Wilkinson's tables (4), the actual probability of two significant results occurring in a series of 25 statistical tests is .36.

The mean figure drawing size of the paranoid group was 6.36 while the nonparanoids produced drawings which averaged 5.76. The *t* test of the difference between means resulted in a value of .98 which is nonsignificant and reveals that paranoids do not typically draw larger figures than other patients.

In a further analysis of the data it was found that the paranoids did not differ substantially from the nonparanoids in the number of indicators occurring in their individual drawings. The mean number of indicators present in the drawings of the paranoid group was 4.97, while the mean for the nonparanoid patients' drawing was 4.23. The *t* test in this instance yielded a nonsignificant value of 1.45.

A final procedure entailed comparing the drawings of three subgroups of paranoid patients categorized according to severity of paranoid symptomatology. The differences between the patients in these groups in the number of indicators appearing in their drawings (drawing scores) were not statistically significant ($F = .24$). The mean numbers of indicators in the three groups were 5.00, 4.41, and 5.40 for the "mildly," "moderately," and "markedly" paranoid patients, respectively.

Summary

An attempt was made to check the usefulness of 26 drawing indicators in determining paranoid pathology. Figure drawings were collected from 61 hospitalized patients of varied diagnostic categories, evaluated specifically for paranoid symptomatology. Comparisons were then made between the figure drawings of 31 patients in this group classified as paranoid and the remaining 30 patients judged entirely free of paranoid characteristics.

Of 25 indicators, emphasis on the outline of the eyes and over-all heavy line emphasis occurred significantly more frequently in the paranoid group. It should be noted, however, that two significant findings in 25 statistical tests could easily be obtained on the basis of chance. The size of the figures did not differentiate between the groups. When the patients were compared with regard to their drawing scores, that is, the number of indicators present in their individual drawings, the groups again did not differ significantly. Nonsignificant results were also obtained in a comparison of the drawing scores of the "mildly," "moderately," and "markedly" paranoid patients.

Received October 4, 1957.

References

1. Holzberg, J. D., & Wexler, M. The validity of human form drawings as a measure of personality deviation. *J. proj. Tech.*, 1950, **14**, 343–361.
2. Machover, Karen. *Personality projection in the drawing of human figures*. Springfield, Ill.: Charles C Thomas, 1949.
3. Ribler, R. I. Diagnostic prediction from emphasis on the eye and the ear in human figure drawings. *J. consult. Psychol.*, 1957, 21, 223–225.
4. Wilkinson, B. A. A statistical consideration in psychological research. *Psychol. Bull.*, 1951, **48**, 156–158.

Journal of Consulting Psychology
Vol. 22, No. 5, 1958

Causative Factors in the Production of Rotations on the Bender-Gestalt Designs

Lewis D. Hannah
Athens State Hospital

Many studies of performance on the Bender-Gestalt Visual Motor Test (1) have been made recently. There are 50 direct references to the test listed in Psychological Abstracts 1950 through 1956. Most of these report correlations of performance on the Bender-Gestalt (BG) with performance on other tests or some manner of behavior. None of them report investigation of reasons as to why various distortions in reproductions are made. It seems that if more of the workings of mental mechanisms tending to cause distortions were known, the test could be used in a more realistic and meaningful way.

The present study is an attempt to show that factors other than mental pathology tend to produce "abnormalities" in the way designs are reproduced. It is believed that the demonstration of even one way of artificially producing an "abnormality" in the reproduction of designs will make necessary re-examination of many premises that have been often taken for granted or even possibly shown by statistical methods to correlate with behavior. Repeated use of the test has suggested that often a "rotation" was produced by a subject as a function of the way the stimulus was presented rather than of something within the subject.

The BG test as commonly used today has the designs printed on a card that is oriented horizontally. That is, the top and bottom edges of the card are longer than the edges on either side. These designs are then copied on paper with horizontal edges which are much shorter than the vertical edges. Because of this, a subject *must* rotate the reproduction 90 degrees in order to preserve the orientation of the design on the card. If this is true, then comparable populations should produce fewer rotations if the stimulus card is presented to them with the design oriented to the short edge of the cards as it now is to the long edge. This hypothesis was tested in the following manner.

First, a new set of BG cards was made. These cards, hereafter referred to as new BG cards in contrast to old BG cards, were made of material similar to the old cards. The design was made to be the same size as on the old cards. The length of the horizontal edges of the cards was also the same as the old cards. However, the vertical edges were extended to a length which made the ratio of the length of the horizontal edge to the vertical edge the same as the ratio of the 8½″ side to the 11″ side of a standard sheet of paper. Sixty consecutive admissions to the Athens State Hospital were tested in the traditional way using the old cards. Then 60 were tested in the same way, except that the stimulus cards used were the set expressly prepared for use in this experiment.

The two groups were then compared for diagnosis, age, sex, etc. By comparing diagnoses it was possible to select 36 cases from one group for which a matching case could be found in the other group. These two groups of 36 each were then compared by Student's t for difference in age and not found to be significantly different. There were 15 males and 20 females in the first group, while there were 12 males and 24 females in the second group.

After the records were selected they were then scored for rotations. For this study, a rotation was scored only if the reproduction was clearly recognizable—if a naive subject could be expected to identify which stimulus card it was copied from—and if it was clearly rotated at least 45 degrees. Each record was then given a score which was equal to the number of designs rotated on the record. Thus the minimum possible score was zero and the maximum possible score was eight. The scores of the records produced by those subjects presented the old cards were then compared with the scores earned by those who had the designs presented on the new cards.

Results

In Group I, those who had copied the designs from the new cards, eight rotations were produced. The average score for the group was .222. The highest score earned was one.

In Group II, the highest score earned was five and the lowest zero. Rotations were found on eight of the records, and the average number of rotations for the group was .639.

Comparison of the scores earned was made by use of Student's *t* and it was found that the probability of the differences in number of rotations produced by the two groups being due to chance alone was less than .01.

Discussion

It is felt that the BG test is a valuable aid in diagnosis but that too often its limitations are not kept in mind when making interpretations. Though too little is known about perceptual mechanisms to explain satisfactorily all "abnormalities" produced by subjects asked to copy the BG designs, this study has demonstrated one way that "abnormalities" are produced as a function of the way the stimulus design is oriented on the card on which it is presented. It is believed that the use of the "new" cards described in this report will eliminate the number of false positives called by the BG. What, in terms of diagnosis or behavior correlates, the positives are, still remains to be established, and interpretations must still be made almost entirely on the basis of personal experience with the test.

Summary

It is believed that factors other than organic or functional impairment of mental functioning will produce rotations in reproductions of the BG cards. To test this, two groups matched for age, sex, and psychological diagnosis were given the BG test. The stimulus cards for the two groups were different in that designs on one were printed with the card oriented horizontally, while for the other test the cards were oriented vertically. The group which was presented the vertical cards produced fewer rotations than the control group which was presented standard BG cards. The results were significant at the .01 level of confidence.

Received September 9, 1957.

Reference

1. Bender, Lauretta. *A visual motor gestalt test and its clinical use.* New York: American Orthopsychiatric Ass., 1938.

Journal of Consulting Psychology
Vol. 22, No. 5, 1958

Sex Differences on the Rorschach[1]

E. Earl Baughman and Samuel Guskin
University of North Carolina

The data to be summarized were gathered as part of a larger study concerned with the role of stimulus attributes in determining Rorschach responses. It was in relation to this larger project that the problem of sex differences arose, since it was not possible to obtain equal numbers of males and females for the study. Numerous investigations in the literature pose an interpretive problem in this regard, although usually the problem tends to be ignored.

Rorschach (9) anticipated that males and females as groups would respond differentially to inkblots even though it was not possible to determine the subject's (*S*'s) sex from the record of the test. Standard references (2, 3, 6, 7, 8, 10, 11) devote little or no attention to this question, however, despite the fact that it is an elementary one in test construction theory. The few empirical studies that have been completed are limited to children and adolescents as *S*s (1, 5, 12). If sex differences do exist, one might expect these differences to show up most clearly in the records of adults. The reason for this is the fact that constitutional differences may be reinforced through cultural learning, producing more clear-cut separation of the sexes as development proceeds.[2]

Procedure and Results

Two examiners (one male, one female) tested a total of 162 *S*s (29 males) following Beck's (3) procedure. Twenty-six males were then selected for whom close matches could be found in the female group as far as the variables of age and intelligence were concerned. Matching selections were carried out by a person unacquainted with the Rorschach protocols. Selections were also made so that each examiner tested equivalent numbers of males and females. All *S*s were home-office employees of two insurance companies in a southern city.[3] The ages of the *S*s ranged between 23–50 ($\bar{x} = 33$ years); Otis IQ scores ranged between 95–135 ($\bar{x} = 111$). In this way two groups of 26 *S*s (one male, one female) were composed which were closely matched on age, intelligence, and place of employment. It was not possible, however, to equate the groups on education, more college males than college females being included in the samples.

The significance of differences between the two groups on 17 summary scores (*R*, *W*, *D*, *Dd*, *S*, *M*, *FM*, *m*, ΣM, *FM*, and *m*, $F+\%$, $T/1R$, T/R, Total Time, $8\text{–}10/R\%$, Number of Content Categories, $A\%$, and *P*) was then evaluated by means of chi square.[4] Only one score, *FM*, revealed a significant difference between the sexes at the .05 level, male *S*s giving more animal movement responses than female *S*s.

In order to check on the stability of this one positive finding, two additional samples of 21 males and 21 females were drawn from a group of 81 *S*s who had responded to an achromatic series of Rorschach blots. Although devoid of color, previous analyses of the data had shown that the frequency of *FM* responses was not affected by the presence or absence of color. These two groups were matched on the productivity (R) variable since precise matching on the other variables previously used for this purpose was not feasible. Analysis of the frequency distributions of *FM* responses by chi square failed to confirm the earlier finding ($p > .30$).

---

[1] This research is part of a larger project supported by Grant M-1027 from the National Institute of Mental Health, U. S. Public Health Service. We wish to thank Hans Grainer and Patricia M. Fossum for administering the tests. Frances Alexander and Paul Fiddleman assisted with the calculations.

[2] It is possible that learning effects may reduce sex differences in a given culture, but in western culture this is not usually held to be the typical outcome of training.

[3] The authors are indebted to the Jefferson Standard Life Insurance Co. and the Pilot Life Insurance Co. for their generous contribution of employee time and testing facilities.

[4] The markedly skewed nature of most distributions of Rorschach scores makes chi square a preferred technique for analysis. The relative insensitivity of chi square, unless samples are large, needs to be recognized however.

Discussion

One other statistical study of Rorschach sex differences has been reported recently. Felzer (4) checked 26 variables, finding two sgnificant differences, *T/R* and *FC*. In his samples of college students, males took more time per response, and they also gave fewer *FC* responses. The females in the present study, in contrast, took longer for each response, although the difference did not reach a statistically significant level. The data do not provide a good check on the color (also shading) variable, since an experimental inquiry procedure was used which renders the data noncomparable; hence, tests for these scores have not been reported and Felzer's findings regarding the *FC* variable must be checked by later work. The data from both studies, however, suggest the improbability that stable differences in summary scores between adult males and females can be demonstrated; an occasional significant difference in one study is not likely to be confirmed by a subsequent study.

There is a broader issue than Rorschach procedure involved in the present discussion; it concerns the relationship between theories of personality and the instruments used to assess personality. Most dynamic theories of personality point out real differences between males and females, related to their very different constitutional endowments, as well as the different types of subsequent critical developmental experiences which they must typically confront. If theory is correct in this regard, then one should expect the organizational properties and dynamics of personality to be distinguishably different in males and females. It does not appear psychologically sound, therefore, to construct personality tests so as to equate the sexes; nor can a given technique, such as the Rorschach, be defended as sufficiently adequate if sex differences cannot be demonstrated.

Summary

Groups of males and females ($N = 52$) were matched for age and intelligence. Rorschach summary scores for the two groups were compared in order to determine if sex differences existed; no stable differences were found. A larger question of the relationship between personality theory and sex differences as related to assessment devices was also discussed. The proposition was defended that sensitive personality devices should reflect sex differences.

Received August 23, 1957.

References

1. Ames, Louise B., Learned, Janet, Metraux, Ruth W., & Walker, R. N. *Child Rorschach responses*. New York: Hoeber, 1952.
2. Beck, S. J. *Rorschach's test. II. A variety of personality pictures*. New York: Grune & Stratton, 1946.
3. Beck, S. J. *Rorschach's test. I. Basic processes*. New York: Grune & Stratton, 1949.
4. Felzer, S. B. A statistical study of sex differences on the Rorschach. *J. proj. Tech.*, 1955, **19**, 382–386.
5. Hertz, Marguerite R. Personality patterns in adolescence as portrayed by the Rorschach ink blot method: I. The movement factors. *J. gen. Psychol.*, 1942, **27**, 119–188.
6. Klopfer, B., & Kelley, D. M. *The Rorschach technique*. Yonkers, N. Y.: World Book, 1946.
7. Klopfer, B., *et al. Developments in the Rorschach technique*. 2 Vols. Yonkers, N. Y.: World Book, 1956.
8. Mons, W. *Principles and practice of the Rorschach personality test*. Philadelphia: Lippincott, 1950.
9. Rorschach, H. *Psychodiagnostics*. Berne: Verlag Hans Huber, 1942.
10. Sarason, S. B. *The clinical interaction*. New York: Harper, 1954.
11. Schafer, R. *Psychoanalytic interpretation in Rorschach testing*. New York: Grune & Stratton, 1954.
12. Stavrianos, Bertha. An investigation of sex differences in children as revealed by the Rorschach method. *Rorschach Res. Exch.*, 1942, **6**, 168–175.

ANNOUNCING A RECORDING OF THERAPEUTIC COUNSELING

The Case of Jim

Provides valuable source material in personality for courses in *personal adjustment* and *general psychology* as well as courses in *counseling* and *psychotherapy*

Contains verbatim passages from a two-year case

Booklets are available for class use; they contain the verbatim material and a section on normal adjustment

The counselor is Dr. Julius Seeman, of George Peabody College

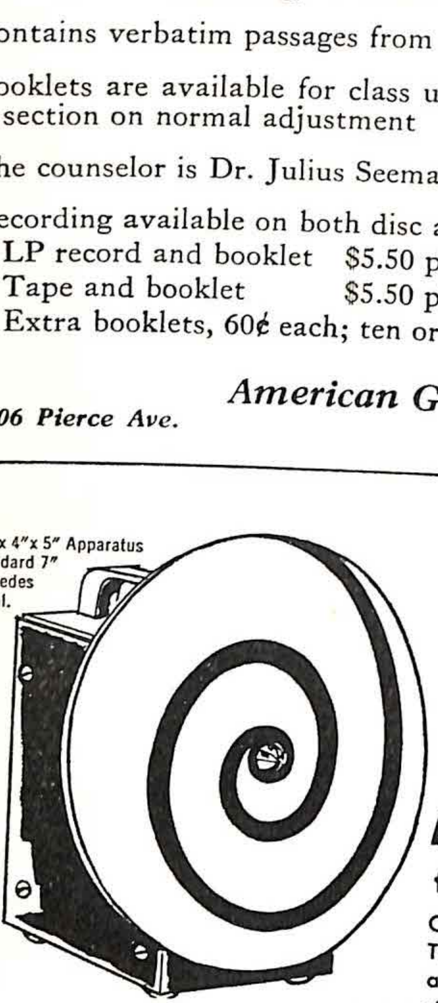
Recording available on both disc and tape

LP record and booklet	\$5.50 plus postage
Tape and booklet	\$5.50 plus postage

Extra booklets, 60¢ each; ten or more, 20% discount

2106 Pierce Ave. ***American Guidance Service, Inc.*** *Nashville, Tenn.*

3"x 4"x 5" Apparatus
with Standard 7"
Archimedes
Spiral.



Spiral

AFTER-EFFECT TEST APPARATUS

for Testing Both Children and Adults

Complete equipment for administering the Spiral After-Effect Test. Recently published research indicates high reliability and validity in discriminating brain injuries from functional states and normals.

Complete with Spiral,
Instruction Manual,
Relevant Clinical Data
and One Year Guarantee

\$47.50 POST PAID

- *Apparatus Is Portable and Compact*
- *Reversible Motor—No Wheel Changing or Turning*
- *Single Switch for Positive and Negative Effect*
- *Built-in "Main" and "Auxiliary" Power Gives Guaranteed Uninterrupted Performance*
- *Uses Regular Flashlight Batteries*
- *Precision Built to Research Specifications*

Order Direct from

PSYCHOLOGICAL RESEARCH & DEVELOPMENT CORPORATION
420 W. LAFAYETTE ST., TAMPA 3, FLORIDA

Journal of Consulting Psychology
Vol. 22, No. 6, 1958

Doctrine and Experience: Their Influence Upon the Psychotherapist

William F. Fey
Department of Psychiatry, University of Wisconsin

The parameters of the psychotherapeutic experience have proved difficult to isolate and quantify. Admitted differences among therapists' behaviors are known to be great (2, 3, 4, 5, 6, 7, 8, 9), and actual differences are doubtless still greater. The most obvious question about psychotherapy—"What happens during the hour?"—we are least equipped with facts to answer. Transcripts of therapy lack the nuances of sound and feeling; recorded tapes neglect important visual clues. Motion pictures require elaborate, unnatural technical settings and are virtually nonexistent; and a panel of qualified but unbiased observers is difficult even to conceive. Moreover, knowledge that the process is being observed or recorded at all is said to inhibit and transmute the experience. The principals themselves, patient and therapist, are least objective of all, for their reports, while firsthand, are inextricably colored by the powerful complex of needs which has produced and sustains their highly unique relationship.

Any simple report of what happens between patient and therapist omits fundamental information since, for example, one's manner in refusing a gift is probably more important in the treatment process than is the fact of refusing it. But the fact and manner are not unrelated, and the patterning of facts ascribable to a therapist doubtless reflects something of his manner and his basic attitudes toward patients. Therefore, despite their obvious limitation as primary data, the present study is based upon therapists' reports of their responses to patients. The assumptions are made that however subtle and refined the underlying rationale of therapeutic strategy may be, it must eventually be given expression in behavior, and that intensive scrutiny of such behavior will afford glimpses of basic therapeutic philosophy. For the present study, then, the working null hypotheses are these: that differences in doctrine and experience among therapists are unrelated both to their direct reports of behavior with patients and to the underlying attitudes which the patterning of these behaviors might suggest.

Method

A simple questionnaire composed of issues commonly arising in the conduct of psychotherapy was sent to each person in a midwestern university city who was known to be conducting consecutive interview treatment. The questionnaire items appear in Table 1. For the first 27 items, the scale of answers ranged from 1 ("yes" or "always" or "routinely") through 5 ("no" or "never"). The final two items summarize the respondents' tendencies in marking the questionnaire items themselves.

The 36 therapists polled included 29 psychiatrists (18 of whom were board-certified), 5 psychologists, and 2 psychiatric social workers. Of these, 34 (94%) responded, one replied to decline participation, and one failed to reply. Each identified his protocol only by stating the number of years he had been doing therapy and by indicating what theoretical orientation, if any, he tended to observe. Doctrine and profession were known to be independent in this sample; the Rogerian group, for example, is not predominantly psychologists. These identifying variables are summarized in Table 2.

Table 1

Questionnaire Items

1. Do you obtain from the patient a detailed case history before beginning psychotherapy (interview therapy)?
2. Would you accept a small gift from a patient in psychotherapy?
3. Would the size of the gift influence your decision?
4. Ordinarily, do you answer general questions of fact such as, "Are psychiatrists always doctors?"
5. As therapist, do you ask your patient questions?
6. Assuming you were licensed, would you write (e.g., sedative) prescriptions for your patient?
7. Do you interview the relatives or friends of a patient, whom you've decided to see in psychotherapy, for the purpose of obtaining more complete knowledge about the patient?
8. Do you ask the patient for dream materials?
9. Do you contact a patient who fails to appear or to notify you on two consecutive appointments?
10. Would you ever go to the patient's home to do psychotherapy?
11. Do you make any efforts to ease the environmental stress upon your patient by contacting significant persons in his life?
12. Ordinarily, do you answer personal questions of fact from the patient such as, "Are you Catholic?" or "Are you married?"
13. Are you willing (if you have the time) to discuss nonemergency issues with your patient on the telephone?
14. Do you deliberately assume different therapeutic roles with different patients?
15. Do you try to maintain the *same* therapeutic role throughout the entire course of therapy with a given patient?
16. Ordinarily, do you answer personal questions of opinion such as, "Do you enjoy classical music?"
17. In the course of therapy your patient decides to make an unpromising marriage. Would you make your doubts about it known to him?
18. Would you deliberately avoid being in a social situation with your patient?
19. If your patient asked you to talk with her husband about her condition, would you be willing?
20. Typically, do you "interpret" your patient's behavior, in the sense of showing him its real significance—meanings of which he is unaware?
21. Assuming that you have open hours, do you offer your patient a choice concerning the number of interviews per week?
22. Do you or would you employ hypnosis or drug preparations to aid your patient in producing material for psychotherapy?
23. Do you extend, beyond the usual time limits, an interview which is proving unusually productive?
24. Do you feel it is important *not* to "personalize" your office with photos, knicknacks, etc.?
25. If a regular patient, not in emergency, phones to request an extra, unscheduled interview, do you grant it?
26. Do you attempt to maintain a definite and consistent length of interview?
27. Ordinarily, would you reply frankly to your patient's question, "Do you think this (therapy) is helping me so far?"
28. How many years have you been doing therapy?
29. The number of midline or "3" answers to the items.
30. The number of extreme or "1" and "5" answers to the items.

Results

Although the sample of respondents well represents a particular locale, serious problems in interpretation arise in that the sample contains no very experienced Rogerians, and the Analytic group is both small and diverse in range of experience. To facilitate the test of hypotheses, the total sample was divided into the following four groups: Rogerians ($N = 7$), Analysts ($N = 5$), Young Eclectics (those 8 therapists with fewer than 10 years' experience), and Older Eclectics (the remaining 14 eclectics with 10 or more years' experience).

The data were first examined with respect to homogeneity within doctrines, testing the expectation that, for example, Rogerians should answer the items more like other Ro-

Table 2

Respondents Grouped by Doctrine and Experience

Years of experience	Doctrines			
	Rogerian	Analytic	Eclectic	Totals
1– 3	5	2	2	9
4– 6	2	1	4	7
7– 9	—	—	2	2
10–12	—	—	4	4
13–15	—	1	4	5
16–18	—	—	—	0
19–21	—	1	3	4
22–24	—	—	3	3
Totals	7	5	22	34

Table 3

Correlations of Each Therapist with the Model or Average of Each Group[a]

Therapist	Models Rog.	Ana.	YEc.	OEc.
Rog. 1	+74	+41	+39	−25
2	+78	+33	+24	−30
3	+82	+30	+24	−42
4	+56	+42	+52	−19
5	+91	+22	+27	−28
6	+72	+18	+04	−16
7	+85	+26	+21	−48
YEc. 1	+28	+13	+32	+09
3	−06	+22	+28	+41
11	+02	+62	+56	+58
16	+48	+37	+42	+03
18	−19	+23	+33	+75
19	+48	+36	+34	−09
20	+60	+47	+31	−36
21	+41	+41	+32	+67
OEc. 2	−21	+23	+03	+48
4	−65	+08	+04	+54
5	−04	+37	+31	+42
6	−32	+07	+32	+80
7	−11	+30	+58	+40
8	+06	+18	+25	+22
9	+22	+31	+35	+31
12	−26	+03	+24	+59
10	−42	+08	+36	+80
13	−49	−20	−17	+43
14	−05	+30	+44	+52
15	−08	+17	+34	+53
17	−27	−20	+16	+46
22	−18	+34	+52	+71
Ana. 1	−25	+27	+34	+24
2	+26	+59	+54	+13
3	+04	+25	+38	+27
4	+70	−12	+04	−41
5	+11	+18	+54	+35

[a] When $df = 27$, $r \gtreqless .37$, $p < .05$; $r \gtreqless .47$, $p < .01$.

gerians than like Eclectics. An informal measure of such homogeneity is to be found in the correlation of each therapist's answers, item by item, with the average response of his own group—himself excluded. These values appear in Table 3, where even a casual inspection would suggest that Rogerians are most homogeneous, the Analysts least. The correlations of Rogerians with their own average have a mean of + .79, Analysts with their average + .25, Young Eclectics + .36, and Older Eclectics + .54.

Each of the four groups has an average response to each item which, when collated, offers a kind of "model" response for that group to the entire questionnaire. It thus becomes possible to compare each therapist with the models of the doctrines he does not claim to represent. These correlations, which also appear in Table 3, further clarify how each therapist relates to the several doctrines examined. Finally, these models may themselves be correlated to produce the figures shown in Table 4, in which it appears that the Analysts and Young Eclectics are most similar as groups in their replies to the items, whereas Older Eclectics and Rogerians are least alike.

Table 4

Correlations Among Group Average Responses[a]

Group	Rogerian	Analytic	Young Eclectic
Analytic	+.36		
Young Eclectic	+.30	+.68	
Older Eclectic	−.36	+.20	+.48

[a] When $df = 27$, $r \gtreqless .37$, $p < .05$; $r \gtreqless .47$, $p < .01$.

Next, a mean score for each group on each item was computed, and these sets of means subjected to separate F tests. The data appear in Table 5, which includes as well the within groups variance ($df = 30$) for readers who may wish to test a private hypothesis regarding item differences. Sharp discrepancies among the group means are clearly present, permitting rejection of the first general hypothesis—that differences in doctrine and experience among psychotherapists are unrelated to their direct reports of behavior with patients.

Finally, as an approach to answering the

Table 5

Differences Among Group Means on Questionnaire Items

Group	Rog.	Ana.	YEc.	OEc.		
N	7	5	8	14	σ_w^2	p
Items						
1	3.6	2.4	2.4	2.1	2.13	—
2	3.4	2.0	2.1	2.8	2.53	—
3	3.4	1.8	2.2	3.6	2.96	—
4	2.9	2.4	2.0	1.3	0.93	.05
5	4.4	2.0	3.2	2.1	1.60	.01
6	4.1	3.2	3.4	3.1	0.87	—
7	4.6	4.2	2.8	2.1	1.03	.01
8	5.0	3.0	3.1	2.4	1.50	.01
9	2.3	3.2	2.9	1.5	2.07	—
10	4.9	3.8	4.6	4.2	1.43	—
11	5.0	3.4	3.0	2.7	0.57	.01
12	3.6	3.2	2.9	1.9	1.47	.05
13	4.9	3.0	2.8	1.9	2.00	.01
14	4.0	3.0	2.8	2.5	3.67	—
15	4.9	2.2	3.5	2.3	2.43	.01
16	4.7	3.4	3.4	2.0	0.67	.01
17	4.9	3.4	3.0	2.4	1.83	.01
18	4.0	3.0	2.8	2.5	2.33	—
19	4.1	2.4	1.4	1.1	0.60	.01
20	4.6	3.6	3.2	2.8	1.03	.01
21	3.6	2.6	1.8	1.9	1.93	—
22	4.6	4.6	4.0	3.3	0.70	.01
23	4.3	3.2	3.1	2.1	1.23	.01
24	3.7	2.2	2.8	1.4	1.63	.01
25	3.0	3.2	2.9	2.1	1.67	.05
26	5.0	5.0	5.0	3.8	1.10	.05
27	5.0	3.6	3.4	2.1	2.33	.01
28	4.7	4.0	4.4	2.6	1.00	.01
29	4.3	3.6	3.4	3.4	1.20	—
30	2.1	3.6	3.5	2.4	1.30	—

second general hypothesis, all 30 items of the questionnaire were intercorrelated, factored by the centroid method, and rotated to yield the oblique solution shown in Table 6.[1, 2] The item communalities are reproduced in this table as evidence of the efficiency of the solution in accounting for the item variances.

The four factors and those items contributing most heavily to them may be regarded as empirically homogeneous yet fairly independent scales in terms of which group differences, if present, might be expected to emerge. To examine this possibility, the purest items from each factor were grouped into scales and each therapist given a score on each scale. Scale I was composed of Items 9, 10, 18, and 30; Scale II of Items 1, 2, 4, 6, and 14; Scale III of Items 7, 11, 22, 26, and 27; and Scale IV of Items 5, 15, and 29. The relative independence of these measures is shown by their intercorrelations in Table 7, and their sensitivity to group differences appears in Table 8.

The clearest test of the effects of doctrine is thought to be between the Rogerian and Young Eclectic groups, and that of the effects of experience between the Young and Older Eclectics. Table 8 suggests that all four scales are sensitive to the variable, Doctrine, and that Scales III and IV are influenced by the variable, Experience.

Discussion

The homogeneity data sustain the impression of Rogerian structure as being relatively well-defined and consistent from patient to patient. Unexpected, however, is the commonality among this group of eclectics who, while avoiding doctrinaire alignments, appear independently to have defined a doctrine of their own. The low agreement among these analysts perhaps reflects mainly their small number and heterogeneous experience.

Comparing individual therapists with their own and with foreign doctrines serves to emphasize the range of behaviors within schools. For example, YEc. 20 and Ana. 4 align with the Rogerians: YEc. 18 and YEc. 21 with the Older Eclectics. Relatively pure partisans appear, such as Rog. 6 and OEc. 4, as well as an eclectic's eclectic (OEc. 9), who has roughly equal loadings in each doctrine.

In choosing the basic groups for further analysis, one is tempted to adopt the func-

---

[1] The writer is grateful to Chester W. Harris of the University of Wisconsin for his counsel on this portion of the study.

[2] A two-page table containing the original correlation matrix has been deposited with the American Documentation Institute. Order Document No. 5738 remitting $1.25 for microfilm or $1.25 for photocopies.

Table 6

Communalities and Item Loadings for Rotated Factorial Solution

Item	Item content	h_j^2	Oblique factors			
			I	II	III	IV
9	No contact after 2 misses	292	−48	31	35	17
10	Goes to patient's home	314	47	00	00	00
18	Does not avoid pt. socially	209	29	01	19	−04
30	Gives many dogmatic answers	552	−52	33	38	−23
1	Takes case history first	411	−26	55	36	14
2	Would accept small gift	417	08	57	03	−15
4	Answers, "Are psychiatrists . . ."	785	−21	79	38	15
6	Writes prescriptions	439	15	44	30	−01
14	Diff. roles with diff. pts.	677	00	74	00	00
21	Sets frequency of interviews	357	−31	39	32	37
7	Interviews relatives	720	−11	43	72	16
11	Eases environment	655	−03	34	67	24
22	Employs hypnosis and drugs	590	−08	34	66	10
26	No fixed interview length	391	00	00	54	00
27	Answers, "Is therapy helping"	661	−04	21	72	34
25	Grants extra hour	320	02	14	54	05
16	Answers, "Enjoy music . . ."	737	−25	35	77	42
24	Personalizes office	709	04	−21	70	45
23	Extends productive hour	702	−18	43	62	46
17	Warns of bad marriage	490	−02	20	58	34
20	"Interprets"	571	37	12	40	08
19	Would talk with spouse	720	17	20	49	47
13	Talks on phone	538	−06	34	51	41
12	Answers, "You married?"	628	−13	49	62	16
8	Asks for dreams	698	−08	45	66	22
5	Asks questions	595	−16	30	35	61
15	Diff. roles with same patient	494	23	−07	27	50
29	Gives many midline answers	431	00	00	00	58
28	Many years doing therapy	616	−24	13	63	63

tional clusters suggested by the correlations of Table 3, rather than the stated allegiances of Table 2. However, such a choice would spuriously maximize latent group differences and serve to deny the fact of significant variations among adherents of a particular school.

The correlations among the four group models shown in Table 4 are thought to reflect some historical developments. The distinction between young and older eclectics is more than one of years. As these data were gathered in 1953, this split corresponds roughly to the pre- and post-World War II training in psychiatry—a period which saw, among other changes, a broader acceptance of modern psychodynamic principles. It thus seems natural that the Young Eclectics should be closer to the Analysts (+ .68) than are the Older Eclectics (+ .20).

Table 7

Correlations Among Scales Based Upon Factor Analysis[a]

Scale	Scale I	Scale II	Scale III
Scale II	+.10		
Scale III	−.00	+.50	
Scale IV	+.21	+.35	+.38

[a] When $df = 32$, $r \gtreqless .35$, $p < .05$; $r \gtreqless .45$, $p < .01$.

Table 8

Group Means, *F* Test Probabilities, and Selected Comparisons Among the Factor Scales

	Group Means					Differences	
Scale	Rogerian	Analytic	Young Eclectic	Older Eclectic	p[a]	Rog. vs. Yng. Ec.	Yng. Ec. vs. Older Ec.
I	16.43	12.00	13.00	14.78	<.07	<.050	NS
II	18.00	13.00	12.62	11.86	<.07	<.050	NS
III	24.14	20.80	18.12	14.00	<.01	<.001	<.025
IV	13.57	7.80	10.12	7.78	<.01	<.025	<.050

[a] For between-within analysis of variance, with 3 and 30 *df.*

Since any intensive search for group differences is certain to disclose them, including spurious ones, it may be wise to underinterpret the individual item variations shown in Table 5. The burden of distilling from these the broader patterns of variation is properly borne by the factor solution. The choice of items for the scales derived from the factor solution was purely empirical, the intent being to use as many items as possible without destroying the relative independence of the scale scores. It evolved that no scales share over 25% common variance, whereas in the original oblique solution, no two factors shared over 13% common variance. While the internal consistencies of scales as brief as these are difficult to establish by traditional methods, such as the Kuder-Richardson formulae, their factorial origin offers direct evidence of their relative homogeneity.

Granting that the interpretation of measures resting upon factored data has always to be provisional, the cues for the hypotheses which follow appear mainly in Table 6. Factor I seems to the writer to represent a dimension involving a distinct, perhaps rigid separation of professional and social roles in the therapist, most pronounced among the Rogerians. One pole of Factor II, with its emphasis upon getting and giving information and help, seems associated with the traditional, rational, flexible approach to health problems, and is least characteristic of the Rogerians. These two would seem to be the "principled" differences among the groups studied.

Factor III is thought to reflect the broad spectrum, resourceful, somewhat supportive approach to patients, and Factor IV, with its focus upon the activity of the therapist, stresses his artful, almost expedient virtuosity in dealing with his patients from moment to moment. Differences in these last two dimensions appear to be associated with both the therapist's doctrine and his experience.

In a series of papers, Fiedler (1, 2, 3) demonstrated that the dimension of communication in psychotherapy (How well does the therapist understand?) is relatively independent of the therapist's doctrine but is sharply associated with his expertness, whereas the dimensions of status (ranging from deference to condescension) and, to a lesser extent, emotional distance (ranging from repulsed to seductive) reflect variations in doctrine but not in expertness. In attempting to relate Fiedler's work to this, it is well to note that, although they often pass as synonymous, experience is not expertness, and on this latter variable the studies may not be directly compared. The relationships among status, emotional distance, and therapeutic doctrine may, however, be examined, provided one is willing to risk some translation of the data.

In general, Fiedler's judges found the Rogerian therapists, irrespective of expertness, to be (perhaps tensely) passive and on equal status with their patients; Analytic therapists were described as aloof. The former finding is perhaps mildly confirmed in the present Items 5, 8, 10, 11, 17, 19, and 20, in all of which the Rogerian therapists shun the initiative. Evidence of analysts' aloofness is not prominent in these data, however, probably since this quality relates almost exclusively to the spirit in which behavior is initiated. It would have been convenient if one of Fiedler's major dimensions could be shown to coincide with certain of the factors isolated by this work; in the writer's judgment, none convincingly does.

Recent work of Strupp (6, 7, 8) offers some basis for comparison. Strupp presented sample situations to therapists, recorded their responses, and then classified these according to the Bales interactional system. Among his conclusions is that with experience comes diversification in technique. The present study tends to confirm this, in that years of experience correlates + .47 ($p < .01$) with number of midline or "3" answers—older therapists entertain more "possible" courses of action. Strupp's conclusion that experienced psychiatrists tend to interpret is not confirmed here, in that Items 20 ("Do you interpret . . .?") and 33 (years doing therapy) correlate only + .18. Strupp also found that, with greater experience, Rogerian psychologists tended to be more like analytically oriented psychologists. While the present study lacks experienced Rogerians as a test of this finding, it is evident from the correlations in Table 4 that experienced eclectics are much less analytically oriented than their younger colleagues.

Some general reservations about the data remain to be stated. The basic measures themselves are primary only in the sense that these are ways in which clinicians are willing to describe themselves. Next, the sample of respondents itself, to the degree it fails to represent the universe of psychotherapists, introduces its own bias into the factorial solution—which solution is, in turn, a permissive method of ordering relationships. Finally, the questionnaire items represent the language and concerns of a clinician whose orientation is largely Rogerian. This fact may in some ways make it more difficult for individuals representing other frameworks to portray fairly and fully their particular manner with patients.

Summary

This study examines the extent to which differences in doctrine and in experience tend to be reflected in the behavior of psychotherapists. The primary data consist of therapists' reports of their handling of issues often arising in treatment. The respondents comprise virtually every psychotherapist from a particular locale, subdivided into these groups: Rogerians, Analysts, Young Eclectics, and Older Eclectics. Correlations among therapists indicate greatest homogeneity among the Rogerians, least among the Analysts. Comparisons of the responses typical for each group suggest that the Analysts and Young Eclectics resemble each other most, while Older Eclectics and Rogerians are least alike. Sharp item differences appear among the four groups studied, and a factor analysis of these items yields four clusters; brief scales based upon these clusters produce dimensions which are differentially sensitive to the variables, doctrine, and experience.

Received November 15, 1957.

References

1. Fiedler, F. E. The concept of the ideal therapeutic relationship. *J. consult. Psychol.*, 1950, **14**, 239–245.
2. Fiedler, F. E. A comparison of therapeutic relationships in psychoanalytic, nondirective and Adlerian therapy. *J. consult. Psychol.*, 1950, **14**, 436–445.
3. Fiedler, F. E. Factor analyses of psychoanalytic, nondirective and Adlerian therapeutic relationships. *J. consult. Psychol.*, 1951, **15**, 32–38.
4. Fromm-Reichmann, Frieda. (Ed.) *Progress in psychotherapy*. New York: Grune & Stratton, 1956.
5. McCary, J. L. (Ed.) *Six approaches to psychotherapy*. New York: Dryden, 1955.
6. Strupp, H. H. The effect of the psychotherapist's personal analysis upon his techniques. *J. consult. Psychol.*, 1955, **19**, 197–205.
7. Strupp, H. H. An objective comparison of Rogerian and psychoanalytic techniques. *J. consult. Psychol.*, 1955, **19**, 1–7.
8. Strupp, H. H. Psychotherapeutic technique, professional affiliation, and experience level. *J. consult. Psychol.*, 1955, **19**, 97–103.
9. Wolff, W. *Contemporary psychotherapists examine themselves*. Springfield, Ill.: Charles C Thomas, 1956.

ournal of Consulting Psychology
/ol. 22, No. 6, 1958

The Problem of Early Termination: Is it Really the Terminee?

Ralph H. Gundlach and Max Geller
Postgraduate Center for Psychotherapy, New York

A recent paper by Taulbee (5) reviewed the question of the early terminator in psychotherapy. The approach was to explore the character structure of this type of patient, as contrasted with that of the patient who continues in therapy. Several studies from other hospitals and clinics on this topic approach the question in similar ways (1; 2, p. 43 ff.; 3; 6).

While these studies report from 30% to 60% of their patients as terminating in 1 to 5 sessions, the rate of termination of patients at the Postgraduate Center, after 1 to 5 sessions, is about 6%.

Data from two surveys at the Center are pertinent. During 1957, 129 patients entered therapy. As of May 1, 1958, 97 of these were still in treatment, while 32 had terminated. Of the latter, 8 had 5 or fewer sessions, and 24 had 19 or fewer sessions. An earlier study had included half of all the cases terminating in 1954. The lowest tally-group of these 130 cases comprised 1 to 10 sessions, which took in 23% of the cases. Of these, 3 had left the city, 2 had been hospitalized, and 2 had died.

What can account for these striking differences? The Center patients seem much like those of any other outpatient clinic in age, education, social status, and problems. Perhaps the screening of patients is somewhat different.

It is our opinion that the difference is largely due to the orientation and objectives of the therapists at the various institutions. We perhaps seek more personality reconstruction than do the therapists at many other institutions (4). Borderline schizophrenics and difficult character disorder patients can get involved in therapy for many months.

Evidence that the Postgraduate Center has a different orientation comes also from data on duration of therapy. In three of the studies cited, it is possible to estimate the median number of sessions: 8, 17, and 26. In contrast, the median number of sessions for patients at the Center terminating in 1954 was about 44. The survey of May 1958 showed 129 who entered therapy during 1957, and 262 others in treatment who had started prior to 1957. Thus, the patient of median duration was in treatment well over one year.

The possibility arises, then, that some of the studies purporting to measure differences in the personality of the "attriters" and "remainers" are also measuring, indirectly, the kind of personality problems that the staff is interested in, or skilled at, handling.

Received August 1, 1958.
Early Publication.

References

1. Cartwright, D. S. Success in psychotherapy as a function of certain actuarial variables. *J. consult. Psychol.*, 1955, **19**, 357–363.
2. Rogers, C. R., & Dymond, Rosalind F. *Psychotherapy and personality change.* Chicago: Univer. Chicago Press, 1954.
3. Rubinstein, E. A., & Lorr, M. A comparison of terminators and remainers in outpatient psychotherapy. *J. clin. Psychol.*, 1956, **12**, 345–349.
4. Strupp, H. H. A multidimensional comparison of therapist activity in analytic and client-centered therapy. *J. consult. Psychol.*, 1957, **21**, 301–308.
5. Taulbee, E. S. Relationship between certain personality variables and continuation in psychotherapy. *J. consult. Psychol.*, 1958, **22**, 83–89.
6. Taylor, J. Relationship of success and length in psychotherapy. *J. consult. Psychol.*, 1956, **20**, 332.

Journal of Consulting Psychology
Vol. 22, No. 6, 1958

Remainer Patient Attributes and Their Relation to Subsequent Improvement in Psychotherapy

Martin M. Katz,[1] Maurice Lorr, and Eli A. Rubinstein
Veterans Administration [2]

Reports from mental hygiene clinics indicate that of all patients accepted for psychotherapy, some 30 to 65% will terminate treatment prematurely against the advice of the therapist. Of those patients who remain, only two-thirds will be subsequently rated improved as a result of treatment (2). These findings have raised a number of questions relative to our understanding of the kind of patient now deriving benefit from psychotherapy. With respect to the problem of predicting length of stay in treatment, Frank et al. (3) have reviewed the significant progress made in this area. In more recent intensive investigation of the problem (5), the general finding that premature terminators and remainers can be distinguished with respect to certain patient attributes, was confirmed. The knowledge, however, that patients who remain do not necessarily improve in treatment led to the question of whether the attributes found to discriminate between terminators and remainers were predictive in some degree of subsequent improvement also. The present study was intended as a preliminary investigation of this problem. In addition, certain therapist variables, such as years of experience, not found to relate significantly to length of stay but presumed to be relevant to the improvement criterion, were included in the study.

Method

As part of a larger study of premature termination of treatment, a nationwide group of 13 Veterans Administration Mental Hygiene Clinics, using a brief self-administered battery of five tests, collected data on all cases considered for psychotherapy (4). Patients examined for hospitalization, neurological disorders, or for brief (half hour or less) irregular treatment were excluded. The intake social worker completed a form at the close of the interview descriptive of such patient characteristics as occupation, annual earnings, highest grade completed, and presenting problems. In addition, the worker made a prediction as to how long the patient would remain in treatment. At the end of the six-month period, if a patient still remained in treatment, the therapist rated the extent to which the patient had changed or improved as a result of psychotherapy.

The 232 cases included in the study were randomly split into two samples of 116 cases. Each subsample consisted of 58 cases who terminated treatment with six weeks of treatment or less and 58 cases with 26 weeks of treatment or more. Analysis of data in the two samples by means of a double cross-validation procedure yielded three short tests which successfully differentiated patients who remain from those who terminate prematurely. One subscale was derived from items included in a Behavior Disturbance Scale, which is designed to elicit information concerning the extent of antisocial behavior in the patient's past, lack of impulse control, lack of personal ties, and lack of goal persistance. Another subscale was selected from the Taylor Manifest Anxiety Scale (6). The third subscale consisted of items selected from the Adorno, Levinson, et al. (1) F scale of authoritarianism. Scores on these three tests and the social worker's prediction of length of stay combined configurally correlated .54 with the terminator-remainer criterion in the cross-validation sample. In addition, annual earnings and the occupational level of the patient were found to discriminate significantly between the terminators and remainers.

---

[1] Now with Psychopharmacology Service, National Institute of Mental Health, Silver Spring, Md.

[2] From the Veterans Administration, Veterans Benefits Office, Washington, D. C.

For the present analysis, three variables descriptive of 116 patients remaining in treatment for six months or longer were correlated with the therapist rating of improvement. These variables were: (*a*) the configural score based on a combination of the three subscales; (*b*) the occupational level of the patient; and (*c*) the patient's annual earnings for the preceding year.

Three therapist variables were also correlated with the rating of improvement, although none had been found to relate significantly to length of stay in treatment. These were included in order to explore in a preliminary fashion the relation of therapists' characteristics to their judgment of improvement. These variables were: (*a*) the therapist's years of experience; (*b*) the therapist's diagnostic classification of the patient; and (*c*) the presence or absence of personal analysis in the experience of the therapist. For the diagnosis, each patient was classified as psychoneurotic, psychosomatic, a character disorder, or as psychotic. In the analyses, the first two and the last two were combined. Product-moment and point-biserial correlations were used.

Results and Discussion

Although all of the above described test and actuarial patient variables are significantly predictive of length of stay in treatment, not one correlates significantly with the measure of improvement employed in this study. Because these results are generally negative, any conclusions that can be drawn from the data must be highly tentative in nature. The findings suggest, however, that knowledge of those factors that determine whether a patient will remain in treatment may have little to tell us about whether he will subsequently improve. The failure of the socioeconomic variables to predict improvement confirms the finding by Frank et al. that patients of lower socioeconomic status, although less likely to remain in treatment, will, if they continue, demonstrate no less improvement than patients from middle economic groups. Consistent with this finding, such configural test patterns as "low authoritarian, low antisocial, high anxiety," found to be highly predictive of continuation in treatment, did not demonstrate any relation to extent of improvement.

Table 1

Correlations of Selected Therapist Variables with Therapist Improvement Rating Criterion in Samples A and B

Variable	Correlation Sample A	Correlation Sample B
Therapist's years of experience	.49*	.39*
Patient diagnostic classification	.34*	.37*

* Significant at .01 level.

Conclusions that can be drawn from these results are somewhat limited also by the nature of the criterion. The therapist's rating, although of importance, can be considered as only one aspect of a complete objective assessment of patient change. An adequate evaluation of patient change would include self appraisals by the patient, objective test data, judgments of independent observers and follow-up data. Consequently, it is possible that the predictors of patient length of stay are related to some other aspects of what is surely a multidimensional criterion.

The results of the analysis of the therapist variables are given in Table 1. The therapist's rating was found to be highly related to his years of experience as a psychotherapist. Lacking other indices of improvement and therapeutic competence, we can only wonder whether increased years of experience actually result in greater success in therapy or whether the standards and perceptions determining improvement ratings of psychotherapists change with increasing years in the field. We do know that the therapist's diagnostic classification of the patient at the end of six months as either (*a*) psychoneurotic or psychosomatic or (*b*) character disorder or psychotic, did not correlate with years of experience, but was highly related to the improvement ratings. This finding suggests that patients who are diagnosed as less severely ill are more likely to be rated improved, regardless of the therapist's experience. Finally, the improvement ratings appeared to have little or no relation to whether the therapist had or had not undergone personal analysis. Obviously more has to be known about the basis of therapist ratings and the relation of such ratings to other more objective indices of improvement.

Summary

The aim of this study was to investigate the relation between patient attributes found to be significantly predictive of length of stay in psychotherapy and subsequent improvement in treatment. A test variable, the configural score, and two actuarial variables (occupational level and annual earnings) were correlated with therapists' ratings of improvement following a six-month period of treatment. No one of the above variables was found to correlate significantly with the criterion. The results suggest that patient attributes associated with subsequent improvement may be very different from those that determine whether a patient will continue in treatment. The nature of the criterion, however, imposes certain limitations on conclusions that can be drawn from the data. It was found, for example, that therapist improvement ratings were significantly related to his years of experience as therapist and to his diagnostic classification of the patient following six months of treatment. The need for investigating additional hypotheses concerning characteristics of potential stayers in psychotherapy in relation to other independent criteria is emphasized.

Received December 2, 1957.

References

1. Adorno, T. W., Frenkel-Brunswik, Else, Levinson, D. J., & Sanford, R. M. *The authoritarian personality*. New York: Harper, 1950.
2. Eysenck, H. J. The effects of psychotherapy: An evaluation. *J. consult. Psychol.*, 1952, **16**, 319–324.
3. Frank, J. D., Gliedman, H. L., Imber, S. D., Nash, E. H., & Stone, A. R. Why patients leave psychotherapy. *A.M.A. Arch. Neurol. Psychiat.*, 1957, **77**, 283–299.
4. Lorr, M., Katz, M. M., & Rubinstein, E. A. The prediction of length of stay in psychotherapy. *J. consult. Psychol.*, 1958, **22**, 321–327.
5. Rubinstein, E. A., & Lorr, M. A comparison of terminators and remainers in outpatient psychotherapy. *J. clin. Psychol.*, 1956, **12**, 345–349.
6. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

Journal of Consulting Psychology
Vol. 22, No. 6, 1958

A Note on McNemar's "On Abbreviated Wechsler-Bellevue Scales"

William Howard
Central Louisiana State Hospital, Pineville, Louisiana

McNemar's article (2), in which he computed the correlations between short forms and the full scale Wechsler-Bellevue for Wechsler's standardization group, contains an error. The culprit is hiding in the first radical in the denominator of the special formula, and reduces all of the reported correlations slightly below their true values.

Although studies have appeared on short-form–full-scale correlations for the WAIS (1, 3, 4), the test which has officially replaced the WB, McNemar's study on the WB remains important since the WB is still being used in some clinical settings. Furthermore, investigators compare their WAIS results with his WB findings.

Maxwell (3) repeated McNemar's procedure on the WAIS standardization group. In general, she found WAIS short forms produce higher correlations than do the WB short forms. Although the best WAIS combinations do not contain the same subtests as the best WB forms, even McNemar's reported WB correlations are lower than the correlations for the same WAIS combinations. Maxwell refers to this in her statement: "With the exception of one duad, all WAIS combinations yield larger correlations than corresponding WB scales." When the WB correlations are corrected however, this last observation is not valid.

The first radical in the denominator for McNemar's general formula is $\sqrt{\Sigma\sigma_j^2 + 2\Sigma r_{ij}\sigma_i\sigma_j}$. In arriving at his special formula he assumes that all σ's are equal to 3. This allows factoring out the σ^2's so that the radical becomes $\sigma\sqrt{n + 2\Sigma r_{ij}}$. The term Σr_{ij}, the sum of subtest intercorrelations given by Wechsler in Table 41 (5) is equal to 20.01, and n is the number of subtests. In the special formula for WB, the value of the radical is $\sqrt{50.02}$, or 7.0725, rather than 7.26 as reported. (The σ cancels out with the same factor in the numerator.) Therefore, all of McNemar's correlations must be increased by a factor of 7.26/7.0725, or 1.0265.

This correction so increases the WB correlations that they exceed the corresponding WAIS correlations for 24 of the 40 short forms reported by McNemar. The WB correlations are equal to the WAIS results in three of the remaining instances.

The correction does not alter the basic result of Maxwell's study. In spite of the increases, the best WB short forms are slightly inferior to the best WAIS combinations. In all cases, but especially for the tetrads and pentads, the differences are very small after the WB correlations are corrected.

Received June 18, 1958.
Early Publication.

References

1. Doppelt, J. Estimating the full scale score on the Wechsler Adult Intelligence Scale from scores on four subtests. *J. consult. Psychol.*, 1956, **20**, 63–66.
2. McNemar, Q. On abbreviated Wechsler-Bellevue scales. *J. consult. Psychol.*, 1950, **14**, 79–81.
3. Maxwell, Eileen. Validities of abbreviated WAIS scales. *J. consult. Psychol.*, 1957, **21**, 121–126.
4. Olin, T. D., & Reznikoff, M. The use of Doppelt's short form of the Wechsler Adult Intelligence Scale with psychiatric patients. *J. consult. Psychol.*, 1957, **21**, 27–28.
5. Wechsler, D. *The measurement of adult intelligence.* (3rd ed.) Baltimore: Williams & Wilkins, 1944.

Journal of Consulting Psychology
Vol. 22, No. 6, 1958

Conditioning of Hostile Verbalizations in a Situation Resembling a Clinical Interview[1]

Arnold H. Buss
University of Pittsburgh

and Ann Durkee
Purdue University

A recurrent obstacle in interviewing and psychotherapy is the verbal inhibition of clients. Topics that are associated with anxiety are avoided or discussed only with the greatest reluctance. Since little progress can be expected unless problems are brought up and discussed, verbal inhibition is clearly an important area for clinical research.

The research may take two forms. First, verbatim interviews could be examined to determine the psychologist's techniques of facilitating verbalization, and then the success of the techniques might be measured. The second form involves setting up a laboratory analogue of the interview situation and investigating the effect of the experimenter's (*E*'s) verbal behavior on the subject's (*S*'s) verbalizations. The present study is of the second kind.

It was felt that while the situation should resemble a clinical interview as closely as possible, discrete trials should be employed for better control of reinforcement by the *E*. Therefore the task of making up sentences was used. This task was first used by Taffel (2), who had a series of 3 x 5 cards, each containing a verb and six pronouns. The *S* was instructed to make up a sentence using the verb and one of the six pronouns. After an initial period of free responding, the *S* was reinforced for every sentence containing the pronoun "I" or "We." Taffel compared this procedure to a therapist's reinforcing a client's talking about himself.

Self-reference content is not usually associated with anxiety, whereas anxiety often accompanies verbalization of hostile content. Often clients are able to talk about their hostile impulses only after their strong inhibitions are removed or weakened. Such inhibition may also be expected in a nonclinical population because it is part of general cultural training to suppress and punish hostility.

The present study seeks to investigate the verbal conditioning of hostile and neutral material. Since inhibition is expected to retard verbalization of hostile content, the learning of hostile material should be slower than the learning of neutral material. Also, since men are allowed a greater variety and intensity of hostile response in our culture, men should learn hostile material faster than women.

Method

Stimuli

Two considerations dictated the initial choice of words. First, the grammatical form should be constant, and therefore only verbs in the past tense were used. Second, slang and pedantry should be avoided, and only relatively familiar verbs were selected.

In compiling a list of hostile words a wide range of intensity was noted, e.g., *stabbed, tortured* vs. *criticized, resented*. It was decided to have two degrees of hostility, mild and intense. A compilation of hostile and neutral words was submitted to six judges,[2] who sorted them into neutral, mildly hostile, and intensely hostile groups. The criterion of acceptance of a word was agreement by five of the six judges.

---

[1] The writers are indebted to Herbert Gerjuoy, George Wischner, and Melvin Manis for their helpful suggestions.

[2] The writers wish to thank Idel Bruckman, Mary Oltean, Jerry Wiggins, and Marvin Zuckerman for serving as judges. The writers served as the fifth and sixth judges.

Next, the Thorndike-Lorge word count (3) was used to match the frequencies of the three types of verbs.[3] Finally, the matched words were placed on 3 x 5 white index cards. There were 30 such cards, each with a neutral, a mildly hostile, and an intensely hostile verb. Since a pilot study revealed that 50 or 60 trials would be required for learning to occur, the number of cards had to be doubled. This was accomplished by using each of the words twice. The second time a word was used, it was accompanied on the card by different words than had accompanied it the first time. For example, *tortured* appeared on a card with *argued* and *invented* the first time and with *flattered* and *rebelled* the second time. The order of placement of the verbs on the cards and the order of presentation of cards were randomized for the series of 60 trials.

Experimental Design

There were two groups of *S*s and 60 trials. During Trials 1–10, the *S*s made up sentences but received no reinforcement. This free responding period established the frequency of response for each type of word prior to conditioning. During Trials 11–60, one group of *S*s was reinforced for using neutral verbs, and the other group was reinforced for using intensely hostile words. There was no group that was reinforced for using mildly hostile words; in this exploratory research the writers felt that any differences in learning should be maximized by using the extreme of hostility. Therefore, of the hostile verbs, only the intensely hostile ones were reinforced, and hereafter the term *hostile verbs* will refer only to intensely hostile verbs.

Procedure

Each *S* was instructed to make up a sentence using one of the three verbs on each card. During Trials 1–10 the *E* said nothing. During Trials 11–60 the *E* said Right after the correct verb was used and Wrong after an incorrect verb was used.

Subjects

The *S*s were 80 college students, 40 men and 40 women. Within each sex, *S*s were assigned randomly to either the neutral reinforced group or the hostile reinforced group, and each group had 20 men and 20 women.

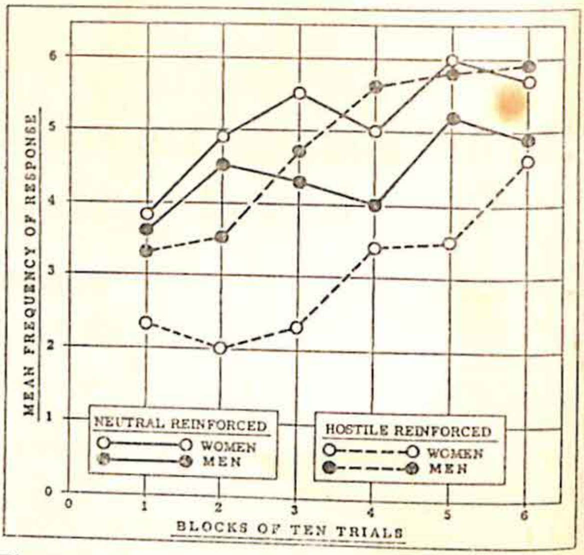


Fig. 1. Acquisition curves of men and women reinforced for hostile and neutral words.

Results

The 60-trial series was divided into six blocks of ten trials each. The measure of learning is the frequency of occurrence of *the response class that was reinforced*. For the Neutral group the frequency of neutral words per block of ten trials is the measure, and for the Hostile group the frequency of hostile words per block of ten trials is the measure. The learning curves for the two groups and for both sexes are plotted in Fig. 1. Inspection of these curves reveals clear-cut differences in the *height* of the curves. Also the *rate* of learning hostile words (the shape of the curves) appears to be slightly faster than the rate of learning neutral words.

The significance of these trends was tested by a Lindquist Type III analysis of variance (1). This analysis of variance is summarized in Table 1. In this table the main effects for Words, the main effects for Sex, and the Words × Sex interaction reflect differences in the *height* of the curves. Differences in the shape of the curves (rate of learning) are reflected in the Words × Trials and Sex × Trials interactions.

The significant F ratios for Words and for Words × Sex indicate that there are signifi-

---

[3] Words were grouped into three categories: 0–24, 25–50, and above 50 times per million according to the Thorndike-Lorge count. Thus the three kinds of verbs were matched on the basis of the frequency *range* in which they fell.

Table 1

Analysis of Variance of Hostile and Neutral Responses of Men and Women for Blocks of 10 Trials

Source	*df*	Mean square	*F*
Words	1	93.64	5.05*
Sex	1	30.00	1.62
Words × Sex	1	200.21	10.79**
Error (between)	76	18.55	
Trials	5	49.45	15.46**
Trials × Words	5	11.47	3.58**
Trials × Sex	5	.66	.21
Trials × Sex × Words	5	4.70	1.47
Error (within)	380	3.20	
Total	479		

* $p < .05$.
** $p < .01$.

cant differences in the height of the four curves. Since there was a significant Words × Sex interaction, four comparisons were made: men's neutral vs. men's hostile; women's neutral vs. women's hostile; men's neutral vs. women's neutral; and men's hostile vs. women's hostile. Two of the four comparisons yielded significant differences. The women's hostile curve was (*a*) significantly lower than the women's neutral curve ($t + 3.19$, $p < .001$ for 76 *df*) and (*b*) significantly lower than the men's hostile curve ($t = 2.63$, $p < .02$ for 380 *df*).

The significant *F* ratio for trials indicates that conditioning did occur (significant increases in frequency of response on successive blocks of trials). The significant Words × Trials interaction suggests that the shape of the neutral curves was significantly different from the shape of the hostile curves. Evidently the rate of learning hostile words is significantly faster than the rate of learning neutral words.

The first 10 trials constituted the free responding period, during which the *E* said nothing after all of the *S*s' responses. Since all *S*s were treated alike during this period, the data from each group may be pooled.

On each trial the *S* could respond with a neutral, a mildly hostile, or an intensely hostile word. In order to simplify the chi square test of difference in frequency of response, the neutral and mildly hostile words were pooled. It was now possible to test the significance of sex differences in frequency of response of intensely hostile vs. neutral and mildly hostile words during the first ten trials.

For the 10-trial period there were 11 possible combinations of frequencies: 10 neutral-mildly hostile, 0 intensely hostile; 9 neutral-mildly hostile, 1 intensely hostile; . . . 0 neutral-mildly hostile, 10 intensely hostile. These combinations for men and women constituted a contingency table which yielded a chi square of 21.2, which is significant at the .01 level of confidence. Thus the women gave significantly fewer intensely hostile and significantly more neutral-mildly hostile responses than the men during the free responding period.

Discussion

It was predicted that neutral material would be learned at a faster rate than hostile material. The results proved otherwise, with hostile words being conditioned faster than neutral words. A variable that was not discussed previously may have contributed to these findings: size of response class. The hostile verbs were sampled from a population of a few hundred hostile verbs, but the population of neutral verbs is many thousands. Learning of a large class of responses is expected to be slower than learning of a small class. Hence neutral verbs were conditioned at a slower rate than hostile verbs, despite the inhibition associated with the latter.

It was also predicted that women would manifest slower learning of hostile verbalizations than men. This prediction was not borne out. Rather, the greater inhibition present in women was revealed in the absolute number of hostile verbs used. Both in free responding and acquisition, women produced significantly fewer intensely hostile verbs than men. The intensely hostile verbs referred mainly to physical violence, e.g., *lynched*, *mauled*, *smashed*, *stabbed*, *strangled*, and *tortured*. The actions represented by these words are severely punished in our culture; they are occasionally allowed by men but never by women. This difference in cultural training would seem to account for the sex difference in frequency of hostile response.

However, the inhibition that depressed frequency of response did not prevent the women from conditioning. While their frequency of hostile response was initially low, it increased with successive acquisition trials. If this finding is confirmed with clinical populations, it would prove encouraging, for it suggests that initial inhibition of response can be overcome by reinforcement. In the investigation of this and other clinical issues the present analogue to a clinical interview would seem to be promising.

Summary

The task of making up sentences was conceptualized as a laboratory analogue of a clinical interview. There were two groups of college students, each group consisting of 20 men and 20 women. One group was reinforced for intensely hostile verbalizations and the other group for neutral verbalizations.

It was found that intensely hostile verbalizations were conditioned faster than neutral verbalizations. This was attributed to the larger size and the diffuseness of the neutral class in comparison to the intensely hostile response class. Women produced significantly fewer intensely hostile responses than men, which was held to be consistent with sex differences in cultural training. Initial inhibition of response did not prevent conditioning from occurring in women. It was suggested that both the findings and the technique have clinical implications.

Received November 25, 1957.

References

1. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
2. Taffel, C. Anxiety and the conditioning of behavior. *J. abnorm. soc. Psychol.,* 1955, **51**, 496–501.
3. Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers Coll., Columbia Univer., 1944.

Journal of Consulting Psychology
Vol. 22, No. 6, 1958

Factor Analysis of Interview Interaction Behavior[1]

Joseph D. Matarazzo, George Saslow
University of Oregon Medical School

and A. Paul Hare
Harvard University

By utilizing an interview standardized along certain minimal dimensions, it has been possible to investigate certain characteristics of the overt interaction patterns of different groups of subjects (8, 9, 10, 11, 12, 13, 14, 15, 16, 17). Results to date suggest that interviewee interaction patterns have the following characteristics. First, there are wide individual differences in interaction patterns among subjects. Second, the interviewee interaction characteristics for any given subject are highly stable across two different interviewers when the latter standardize their interviewing behavior along the minimal predefined dimensions of our standardized method, and, at the same time, these interviewee characteristics are modifiable by planned changes in the intra-interview behavior of either interviewer (10, 15, 17). Third, the marked stability and modifiability of an individual's interaction patterns, found for our first sample of subjects (15), were cross-validated in a second sample (9). And fourth, the stability and modifiability were equally striking when only a single interviewer was used and the test-retest interval was extended to seven days (16), five weeks (17), and eight months (17), in contrast to the first two studies which employed a test-retest interval of a few minutes.

The finding that interviewee interaction patterns are stable under identical (along certain dimensions) interpersonal environmental stimulus conditions and modifiable when these environmental conditions change can be thought of as constituting one type of evidence of the construct validity of the interaction patterns (1, pp. 13–28; 5, p. 288). However, most of our research to date has concerned itself with establishing the reliability of these interview interaction variables. Thus we have studied and established as having extremely high reliability the following aspects of interview interaction behavior: the reliability of the *interviewer* who serves as the independent variable by conducting the standardized interview (9, 15, 17); the reliability of the *interviewee* interaction patterns, the dependent variables (9, 15, 16, 17); the reliability of the *observer* who observes the interviewer-interviewee interaction and records his observations by pressing separate keys for each participant (14); and, finally, the reliability of the *scorer* who scores the final Interaction Chronograph record (15, p. 429).

Having thus established unusually high reliabilities for these interview interaction variables, it was appropriate to turn to the question of their validity, or "meaning." It was apparent to us that the question of the validity of our measures, as of most personality measures derived from other frameworks, was made all the more difficult to investigate by the fact that there exists no well-defined or highly integrated interaction theory of personality from which hypotheses relevant to validity could be generated. For this reason, we have approached the validity problem from a number of points of view; amounting largely to what Cronbach and Meehl call the "bootstraps" method (5, pp. 286–287). When

---

[1] This investigation was supported by a research grant (M-1938) from the National Institute of Mental Health, of the National Institutes of Health, U. S. Public Health Service.

sufficient studies bearing on validity have been conducted and reported, it is hoped that a preliminary theory of interaction can be formulated and new, specific hypotheses generated and tested. To date we have reported one validity study (12) and soon will be reporting two additional ones (11, 13).

The work here being reported, a factor-analysis of the 12 interview interaction variables used by us before the recent addition of two additional variables, is an example of what Cronbach and Meehl call construct validity (5, p. 287). The three other validity studies are more properly examples of concurrent validity (1). The first, already reported, deals with psychological test correlates of these interview interaction behaviors (12). The second, on the relationship between some aspects of the verbal content of the interview and these formal overt interview interaction patterns (13), and the third on differences in interview interaction patterns among five different psychodiagnostic groups (11), are now in preparation.

Procedure

The data which were factor-analyzed for the present study consisted of the standardized interview interaction scores of 60 subjects (10). The *S*s were unselected and were interviewed in their order of referral to the psychiatric outpatient clinic of a large, urban medical center. There were 32 females and 28 males, all white, ranging in age from 16 to 62 (median age of 33), who had presenting complaints typical of outpatients at most psychiatric facilities. The standardized interview was conducted by two experienced interviewers. For purposes of establishing reliability of the measures, each *S* was interviewed twice. For 40 of the *S*s, the two interviews were conducted during the same afternoon, usually with a 2- to 5-minute interval between interviews, and in counterbalanced order by the two interviewers. All of the test-retest interviews with the remaining 20 *S*s of the present study were seven days apart and were conducted, on both occasions, by only one of these two interviewers (the psychiatrist).

The Interaction Chronograph, devised by Chapple (2, 3, 4) for recording certain temporal aspects of verbal and gestural interaction, was the instrument of measurement used during the 60 interviews. The rules of the standardized interview (8) require the interviewer to follow certain simple predefined behavioral patterns (for example, that each of the interviewer's comments should be 5 seconds in duration). In addition, the rules require the interviewer to vary his behavior somewhat during each of five predefined subperiods of the interview. The interviewer's varied behavior during these five periods constitutes a sequence of different behavioral situations, or what can be thought of as varying miniature interpersonal situations, designed to elicit the interviewee's characteristic interaction patterns under different interpersonal conditions. Periods 1, 3, and 5 consist of *free* give-and-take interviewing, while Periods 2 (silence) and 4 (interruption) are *stress* phases of the interview (8). It should be noted that the stress of Periods 2 and 4 is provided by the overt behavior ("not responding" and "interrupting," respectively) of the interviewer, and not by the "content," or what he says to the interviewee. The durations of Periods 1, 3, and 5 are *fixed* (10, 5, and 5 minutes, respectively), whereas the durations of Periods 2 and 4 are *variable*. The latter periods consist, respectively, of 12 failures to respond or 15 minutes, whichever is shorter, and 12 interruptions by the interviewer or 15 minutes, whichever is shorter. The length of the standardized interview ranged for these 60 *S*s from 24.5 minutes with one patient to the full fifty (50.3) minutes for another. The mean length of the interview was 32 minutes with this outpatient sample.

More complete definitions of the interaction variables here factor analyzed can be found in our earlier reports (8, 15). Briefly, they are: (*a*) *Pt.'s Units*, the number of times the patient acted; (*b*) *Pt.'s Action*, the average duration of the patient's actions; (*c*) *Pt.'s Silence*, the average duration of the patient's silences; (*d*) *Pt.'s Tempo*, the average duration of each action plus its following inaction, as a single measure; (*e*) *Pt.'s Activity*, the average duration of each action minus its following inaction, as a single measure; (*f*) *Pt.'s Adjustment*, the durations of the patient's interruptions minus the durations of his failures to respond, divided by Pt.'s Units; (*g*) *Dr.'s Adjustment*, the durations of the interviewer's interruptions minus the durations of his failures to respond, divided by Pt.'s Units; (*h*) *Pt.'s Synchronization*, the number of times the patient interrupted or failed to respond to the doctor, divided by the number of Pt.'s Units; (*i*) *Dr.'s Units*, the number of times the doctor acted; (*j*) *Pt.'s Initiative*, the percentage of times, out of the available number of opportunities (usually 12) in Period 2, in which the patient acted again (within a 15-second limit) following his own last action; (*k*) *Pt.'s Dominance*, the number of times in Period 4 that the patient "talked down" the doctor minus the number of times the doctor "talked down" the patient, divided by the number of Pt.'s Units in that period; (*l*) *Pt.'s Quickness*, the average length of time in Period 2 that the patient waited before taking the initiative following his own last action.

Table 1

Intercorrelations (Pearson *r*'s) Among Interaction Variables Based on Interviews in a Sample of 60 Subjects; *First* Interviews Above the Diagonal, *Second* Interviews Below the Diagonal

Variables	1	2	3	4	5	6	7	8	9	10	11	12
1. Pt.'s Units		−73	17	−82	−83	−20	−05	61	97	−21	01	−29
2. Pt.'s Action	−73		−30	92	88	22	20	−54	−71	25	12	37
3. Pt.'s Silence	19	−27		−22	−41	−76	−73	55	23	−68	−52	−84
4. Pt.'s Tempo	−83	95	−19		96	19	22	−52	−79	17	05	35
5. Pt.'s Activity	−81	95	−36	98		34	33	−63	−82	30	13	51
6. Pt.'s Adjustment	−08	19	−78	11	26		48	−45	−25	50	44	61
7. Dr.'s Adjustment	03	32	−76	22	35	50		−33	−07	57	46	87
8. Pt.'s Synchronization	64	−36	47	−42	−47	−26	−08		70	−74	−21	−64
9. Dr.'s Units	95	−62	25	−71	−70	−13	06	74		−28	06	−34
10. Pt.'s Initiative (Per. 2)	−33	36	−62	33	42	32	59	−58	−39		33	76
11. Pt.'s Dominance (Per. 4)	−03	17	−32	17	22	37	33	−01	12	05		43
12. Pt.'s Quickness (Per. 2)	−30	41	−85	35	49	49	85	−54	−33	80	23	

Note.—For $r = .25$, $p < .05$.
For $r = .32$, $p < .01$.

Results

The retest standardized interview on each of the 60 *Ss* made it possible to replicate and thus check the reproducibility (through an attempt at cross-validation) of the results of the factor analysis of the 60 first interviews. In Table 1 is shown the matrix of intercorrelations (Pearson *r*'s)[2] of the 12 interview interaction variables derived from the 60 interviews. The intercorrelations for both the first and second (replication) interviews are included in the one table in order to facilitate comparison. Intercorrelations for the data of the first interview are given above the diagonal in Table 1, while the intercorrelations for the second interview are shown below the diagonal. As can be seen, the intercorrelation values for the replication interview are strikingly similar to those shown for the original interview and constitute a clear cross-validation of the findings with the 60 first interviews.

Before going on to the factor analysis of these data, several points in Table 1 are worth noting. The first of these is the high negative correlation (−.73) between how often a patient speaks (Pt.'s Units) in the standardized interview and the average duration of each of these units of utterance (Pt.'s Action). It would appear that people who speak often in the standardized interview do so in relatively brief utterances, whereas people who speak infrequently tend to do so in correspondingly longer utterances. On the other hand, since a correlation between two events merely indicates an existing relationship but tells us nothing about which event influences the other, one can also interpret this −.73 relationship to mean that people who habitually speak in brief utterances do so relatively frequently; whereas individuals whose utterances tend on the average to be quite lengthy, seem to speak much less often. This finding would indeed be an interesting addition to our knowledge of human behavior were it not for the possibility that it may represent an artifact of the standardized interview. It will be remembered that three periods [1, 3, and 5], constituting 20 minutes of the standardized interview, are of fixed duration, whereas, on the average, only 12 minutes (but potentially 30 minutes for all *Ss*) are of variable length.[3] Thus, in a partially "closed system," whose limits on durations (and thus indirectly on units) are prescribed in advance, it is not surprising that a negative correlation should exist between number of units and average duration of action. It is clear that an *S* who acts 112 times during a 32-minute interview must do so in briefer durations than a second *S* who interacts only 25 times during the same period. The high negative correlation is probably not all artifact, however, since up to 30 minutes of the interview are variable and unfixed in regard to duration, with the interviewer terminating each of Periods 2 and 4 after anywhere between 1 and 15 minutes, depending on the patient's "unrestricted," in this sense, behavior in these periods. However, it is possible better to test the degree of correlation between number of units and average duration of action by observing a series of pairs of two individuals in a free conversation, or a single pair of individuals in a series of free conversations (free of predetermined time limits). We plan to carry out such studies as a more adequate test of this initial finding.

---

[2] The frequency distribution of the interaction scores of 134 *Ss*, representing a normal population as well as several diagnostic groups, sufficiently approximated a bell shape to meet the required statistical assumption of a normal distribution. Pt.'s Initiative deviated somewhat, with some tailing-off at the low end (0.00) and piling-up at the high end (1.00).

[3] For the 60 *Ss* of this study, the range for the *combined length of the two variable periods* (Periods 2 and 4) was from 4 minutes to 30 minutes, with a mean of 12 minutes.

A second finding of interest in Table 1 is the relative independence of Pt.'s Units and the following variables: Pt.'s Silence; Pt.'s Adjustment; Dr.'s Adjustment; Pt.'s Initiative; Pt.'s Dominance; and Pt.'s Quickness. In part, this is in accordance with Chapple's original plan since, as with a technique like the Rorschach, in order to make comparable the results from *S*s with different numbers of interaction units, all Interaction Chronograph variables are divided by the number of units for that individual (8, p. 354). It is clear, however, that even though the influence upon *all* variables of different numbers of units among different *S*s has been thus partly controlled, there is still a highly significant negative correlation between number of units and one of these variables, average duration of action (described above). That is, after relating all variables to the same scale, Pt.'s Units still correlate with Pt.'s Action, whereas Pt.'s Units does not correlate materially with Pt.'s Silence, Pt.'s Adjustment, Pt.'s Initiative, Pt.'s Dominance, and Pt.'s Quickness. Thus, in descriptive language it would appear that how long a person is silent (in surprising contrast to how long he acts); how he maladjusts, when he does so in two-person interactions; how often he takes the initiative; and how quick he is to do so; and how often he dominates or is dominated are all behavioral characteristics which are independent of his total number of units (i.e., his frequency of interaction).

Other findings of interest seen in the first interview data in Table 1 are: the moderate negative correlation between action and silence (−.30); the moderate positive correlation between duration of action and frequency of initiative-taking (.25); the independence of duration of action and frequency of dominance (.12); the high negative correlation between silence and initiative (−.68), i.e., the person with the longest silences shows the fewest initiatives and vice versa; the positive correlation between adjustment and initiative (.50), and between adjustment and dominance (.44), and between adjustment and quickness (.61); the moderate positive correlation between Pt.'s Initiative and Pt.'s Dominance (.33), which, however, was not cross-validated (.05); and, finally, the high correlation between how often one takes the initiative in Period 2 and how quick he is to do it (.76).

Both sets of intercorrelations shown in Table 1 were factor analyzed, the factors being extracted by using Thurstone's centroid method[4] (18, pp. 161–170). In Table 2 are shown the *original centroid* factor loadings and the *rotated* factor loadings for the 60 first interviews, while in Table 3 are shown similar values for the sample of 60 second (replication) interviews.

From an inspection of Tables 2 and 3 it is clear that the results of the factor analysis of the 60 first interviews were highly reliably cross-validated on the sample of 60 second interviews. Thus, it is necessary only to describe the results shown in Table 2. From the column of rotated factor loadings under Factor I in this table it would appear that this first factor is in part a latency, or speed

---

[4] In extracting the factors, using Thurstone's centroid method as a base, computations were made on IBM machines, using procedures developed by Arthur Couch with assistance from Eugene Kassbaum, both of the Department of Social Relations at Harvard University.

Table 2

Factors for Interaction Variables Based on *First* Interview of 60 Patients

Variables	Centroid Loadings					Rotated Loadings				
	I	II	III	IV	V	I	II	III	IV	h²
1. Pt.'s Units	70	−63	17	11	14	05	−94	20	−04	93
2. Pt.'s Action	−74	49	25	04	16	−18	87	20	−15	85
3. Pt.'s Silence	75	56	08	13	−07	93	−13	15	03	90
4. Pt.'s Tempo	−74	57	30	−02	04	−12	93	20	−23	96
5. Pt.'s Activity	−85	46	20	−09	04	−28	93	08	−20	98
6. Pt.'s Adjustment	−62	−41	−18	−33	10	−73	15	−36	−11	69
7. Dr.'s Adjustment	−62	−55	30	12	−16	−83	05	30	−13	79
8. Pt.'s Synchronization	79	−07	38	−26	−10	51	−61	08	−45	84
9. Dr.'s Units	72	−60	29	05	14	08	−93	24	−17	96
10. Pt.'s Initiative (Per. 2)	−66	−38	−30	41	09	−74	20	08	50	84
11. Pt.'s Dominance (Per. 4)	−37	−44	11	−14	10	−57	−05	−02	−18	36
12. Pt.'s Quickness (Per. 2)	−82	−47	06	19	−14	−91	25	18	09	93
						Sil.	Act.	Adj.	Init.	

Table 3

Factors for Interaction Variables Based on *Second* Interview of 60 Patients

Variables	Centroid Loadings					Rotated Loadings				
	I	II	III	IV	V	I	II	III	IV	h²
1. Pt.'s Units	72	−63	14	12	07	06	−95	18	−01	95
2. Pt.'s Action	−78	40	37	17	−07	−27	83	38	−14	93
3. Pt.'s Silence	73	59	13	19	16	93	−10	23	04	93
4. Pt.'s Tempo	−77	52	33	12	−04	−18	91	32	−15	99
5. Pt.'s Activity	−87	39	33	07	03	−34	89	28	−18	102
6. Pt.'s Adjustment	−53	−48	07	−41	−06	−71	03	−24	−34	68
7. Dr.'s Adjustment	−59	−65	21	32	−21	−88	−04	37	08	92
8. Pt.'s Synchronization	66	−16	51	13	−18	35	−58	45	−27	74
9. Dr.'s Units	69	−58	38	13	19	08	−90	36	−18	97
10. Pt.'s Initiative (Per. 2)	−69	−24	−34	34	16	−66	32	00	48	76
11. Pt.'s Dominance (Per. 4)	−27	−27	36	−23	09	−38	00	09	−42	33
12. Pt.'s Quickness (Per. 2)	−80	−47	−22	30	−05	−90	23	06	37	100
						Sil.	Act.	Adj.	Init.	

of response measure. That is, it reflects how long a patient is silent, on the average, before he responds to the interviewer's last utterance. For simplicity one can label this a "silence" factor. Patient interview interaction variables having the highest rotated loadings on this first factor are Pt.'s Silence (.93), Pt.'s Adjustment (−.73), Pt's Synchronization (.51), Pt.'s Initiative (−.74), Pt.'s Dominance (−.57), and Pt.'s Quickness (−.91). By Interaction Chronograph measurement, Pt.'s Silence and Pt.'s Quickness are both measures of a patient's average duration of interview silence: the former variable measures this for all five periods, the latter variable only during Period 2. The minus sign of the factor loading for Pt.'s Quickness (−.91) should be disregarded in the interpretation of this first factor as the negative direction is an artifact of measurement due solely to the inverse manner in which this variable is recorded in the present model of the Interaction Chronograph. (One or two other variables pose similar but unimportant "sign" artifacts.) Pt.'s Adjustment, which has the next highest loading on Factor I, is also a silence measure, i.e., the average of the durations of a patient's interruptions minus the durations of his failures to respond (silences).

The last two variables with loadings on this factor, Pt.'s Initiative and Pt.'s Dominance, are both frequency measures and thus differ from the former, all of which are duration measures. Pt.'s Initiative represents the percentage of times the patient took the initiative (i.e., spoke again following his own last utterance) during the interviewer's 12 prescribed failures to respond in Period 2. Thus, in a sense, it might be speculated that the Pt.'s Initiative variable could reflect a "drive"[5] aspect of human interaction (i.e., an additional parameter of an individual's frequency of interaction) or, at the least, more than what is measured by the duration measure, Pt.'s Silence, alone. That is, how long one waits before he acts (Pt.'s Silence) as an Interaction Chronograph measure, would seem not to reflect all that is involved in the Pt.'s Initiative variable and vice versa. For there are some patients who will act, i.e., will take the initiative, and there are others who will not under conditions of interviewer silence. Both types of patients receive durations of silence scores, while only the former group receives an initiative score for this unit of interaction. (The tenability of this dual-nature hypothesis of Pt.'s Initiative is strengthened by the fact that Pt.'s Initiative is the best measure of the Fourth Factor, with a factor loading of .50 on it. This will be described below.)

The factor loading of $-.57$ for Pt.'s Dominance on Factor I indicates that there may be still a third parameter present in this "duration of silence" first factor. That is, this factor loading of $-.57$ for Pt.'s Dominance on Factor I ("silence") shown in Table 2, and the r of $-.52$ between Pt.'s Dominance and Pt.'s Silence shown in Table 1, indicate that people who tend to have long durations of silence during their interview tend also significantly not to dominate or out-talk the interviewer when he interrupts them 12 times in Period 4. This finding that patients with longer average interview silences exhibit a greater frequency of terminations of speech (submissions) when interrupted, would seem to open up a new area for future research to further our understanding of still other aspects of two-person interactions.

The second rotated factor shown in Table 2 (and cross-validated in Table 3) appears to reflect "duration of patient action per unit," i.e., how long on the average the patient spoke or otherwise communicated during each of his interaction units. This "action" factor reflects the earlier discussed high negative correlation between Pt.'s Units and Pt.'s Action, as can be seen by the factor loading of $-.94$ for Pt.'s Units and of .87 for Pt.'s Action on Factor II. Other patient variables with high loadings on the second factor are: Tempo and Activity (both .93), and Synchronization ($-.61$). The loading of $-.93$ for Dr.'s Units on this factor reflects the r's of .97 and .95 (shown in Table 1, above and below the diagonal, respectively) between Pt.'s Units and Dr.'s Units. The latter near unity correlations are not surprising since the rules of the partially standardized interview require the interviewer (in all periods but the second) to take his lead from the interviewee and to respond with a five-second utterance to each of the patient's communication units, thus earning the interviewer one Dr.'s Unit for each Pt.'s Unit.

The loadings on Factor II of .93 for Pt.'s Tempo and .93 for Pt.'s Activity are interesting since they indicate that the "action" factor is highly saturated with these two derived interview interaction variables. Since these same variables do not appear in the "silence," or first factor (loadings of $-.12$ and $-.28$, respectively), it would appear that, as interview variables, Pt.'s Tempo and Pt.'s Activity are made up primarily of Pt.'s Action rather than Pt.'s Silence; despite the fact that Pt.'s Tempo is scored as the average sum of Pt.'s Action plus Pt.'s Silence while Pt.'s Activity is scored as the average Pt.'s Action minus Pt.'s Silence. This fact is shown even more clearly in Table 1, where it is seen that Pt.'s Action has very high correlations with Pt.'s Tempo and Pt.'s Activity (.92 and .88, respectively, in the first interview; and .95 and .95 in the replication interview) whereas Pt.'s Silence has only minimal correlation with these same variables (−.22 and −.41, respectively; and −.19 and −.36). In part this is due to the restricted range of durations of silence (from 5 to 19 hundredths of a minute for this sample) relative to the greater range of durations of action (from 11 to 257 hundredths of a minute) (10).

---

[5] Some evidence for this hypothesis (12, p. 332) comes from a cross-validated statistically significant r of .44 between our Pt.'s Initiative variable and a patient's score on the Taylor Anxiety Scale (purported to be a measure of Hullian drive level).

Thus the high loadings of Pt.'s Tempo and Pt.'s Activity, as well as that of Dr.'s Units on Factor II, would appear to be reflections of the derived nature of these two former variables, including the differences in range for silence and action, and the standardized rules governing the interviewer's behavior in regard to the Dr.'s Units variable. With the exception of Pt.'s Synchronization, which is highly correlated with Pt.'s Units (see Table 1), the absence of significant loadings of any of the other interview variables on Factor II would indicate that it is purely a reflection of how often a patient communicates (Pt.'s Units) and how long on the average he talks once he has begun (Pt.'s Action). Although Factor II justifiably could thus equally be thought of as reflecting units, or one's number of interactions (a frequency measure), we prefer for the time being to call it "action" and thus think of it as a duration measure. Only further research will permit us to begin to unravel the interesting question of whether one's frequency of (inter)action determines his duration of action, or vice versa, or neither (as, for example, in free in contrast to standardized interviewing).

Rotated Factor III in Table 2 appears to be a much weaker factor than either Factors I or II; with no factor loading reaching a level above (minus) .36. This weak factor is saturated most heavily with Pt.'s Adjustment (factor loading of −.36), and less so by Pt.'s Units (.20), Pt.'s Action (.20), Pt.'s Tempo (.20), Dr.'s Adjustment (.30), Dr.'s Units (.24), and Pt.'s Quickness (.18). These results are cross-validated for the most part in Table 3, although, in this table, Pt.'s Synchronization (the frequency with which the patient maladjusts, which with the typical subject involves a frequency count of how often he fails to respond quickly to the interviewer rather than how often he interrupts the interviewer) has the highest factor loading (.45). Since all the other interview variables with loadings on Factor III in Tables 2 and 3 reflect a duration measure in one way or another—either of actions or silences—and since Pt.'s Adjustment has one of the highest factor loadings, we tentatively will consider this factor to reflect one's "average duration of (mal)adjustments"; remembering, however, that the frequency of these maladjustments (i.e., Pt.'s Synchronization), also may be reflected in this third factor. That Factor III is neither well established nor strongly present and thus is a weak factor, however, is also clear from the results.[6]

In Table 2 Rotated Factor IV seems to be most heavily saturated with Pt.'s Initiative (factor loading of .50), although Pt.'s Synchronization has a loading of −.45 on this same factor. On replication (Table 3), the factor-loading for Pt.'s Initiative is cross-validated (.48), while that for Pt.'s Synchronization is only moderately so (−.27). Pt.'s Dominance, which in Table 2 had an insignificant loading of −.18 on this fourth factor, shows a significant loading of −.42 during the second, or retest interview (Table 3). However, since the test-retest reliability for measurement of the Pt.'s Dominance variable is only moderate (9, p. 273; 14, pp. 272–273), in contrast to the very high reliability for all the other variables, the full meaning of this finding remains unclear. Thus, in view of the fact that Pt.'s Initiative had the highest loading on Factor IV during the first interview (.50) and this was cross-validated in the second interview (.48), while the significant loadings for the other variables were not cross-validated by appearing in both interviews, we tentatively will consider that the fourth factor reflects primarily the habitual degree of "initiative" which an *S* will manifest under conditions of planned interviewer silence.

---

[6] Further evidence for the tentative nature of Factor III (and Factor IV) comes from a second factor analysis of the data in Table 1, done for us by W. A. Botzum and Charlotte David of the University of Portland. Using an oblique solution in contrast to the orthogonal solution shown in Tables 2 and 3, their analyses confirmed our Factors I and II, but indicated that Dr.'s Adjustment and Pt.'s Quickness had the highest consistent loadings on Factor III (the high Factor loading for Dr.'s Adjustment on Factor III is also present in Tables 2 and 3), and that, whereas Pt.'s Initiative had a high loading on Factor IV in the first interview, this was not cross-validated in the data from the second interview.

In the last column of Table 2 (and Table 3) are shown the communalities (h^2) of each interview interaction variable. The values in this column represent the amount (proportion) of the variance of each interview variable which is accounted for by our four factors (6, p. 492). It is clear from Table 2 (and cross-validated in Table 3) that the four factors which have emerged from the factor analysis (pt.'s "silence" behavior, pt.'s "action" behavior, pt.'s "adjustment" behavior, and pt.'s "initiative" behavior) exhaust almost the total variance of the following interview variables: Pt.'s Units (93%),[7] Pt.'s Action (85%), Pt.'s Silence (90%), Pt.'s Tempo (96%), Pt.'s Activity (98%), Dr.'s Units (96%), and Pt.'s Quickness (93%). They account for the variance of Dr.'s Adjustment (79%), Pt.'s Synchronization (84%) and Pt.'s Initiative (84%) to a slightly smaller extent. The variances of two interview variables are not materially accounted for by these four rotated factors. They are Pt.'s Adjustment (69%), which most helps to define our weak third factor, and Pt.'s Dominance (36%) which, with a cross-validated moderate factor loading on Factor I, and a noncross-validated loading on Factor IV, has only about one-third of its variance accounted for by the four factors. In view of these findings it seems safe to conclude that, on the one hand, more will have to be learned about the Pt.'s Adjustment variable and, on the other hand, as suggested by us in an earlier publication (14, p. 272), considerable error variance still exists in the measurement of our Pt.'s Dominance variable. Refinements in definition, or increase in the length of the interruption subperiod (for example, lengthening it to a *fixed* 15 minutes rather than leaving it free as now, with an average duration of only approximately two to four minutes) may permit us to reduce the aforementioned error variance of Pt.'s Dominance and thus learn more about its nature.

By squaring each row entry in each of the four columns of rotated factor loadings in Tables 2 and 3 and then summing down the columns for each of these four rotated factors, one can obtain an estimate of how much of the total variance is accounted for by each factor (6, p. 492). Rounding off, the four factors named at the bottom of each column account individually for 41%, 47%, 5%, and 7% of the total variance, respectively, in Table 2, and 38%, 44%, 9%, and 9%, respectively, in Table 3. It is thus clear that, in terms of the Interaction Chronograph variables so far employed by us, an *S*'s interview interaction behavior consists primarily of actions and silences; and moderately of his pattern of adjustment (and synchronization) with the interviewer, on the one hand, and his degree of initiative-taking when the interviewer attempts to assess this by purposely remaining silent, on the other. To date we have not worked with Chapple's temperament and personality variables (4). Although these represent various arithmetic combinations or ratios of the 12 variables shown in Table 1 of the present paper they, nevertheless, still may contribute additional factors other than the four reported in the present paper.

Discussion

The results of the present study may be compared to the findings reported by Lundy in an unpublished study (7). By employing two therapists with one patient for 38 interviews (each therapist saw the patient individually for 19 interviews in an odd-even design), Lundy was able to relate interview Interaction Chronograph variables to certain *clinical impressions regarding the progress of psychotherapy* as these qualitative data were assessed from the typewritten interview transcripts. In addition to demonstrating some interesting relationships of this type in a study which used "free" psychotherapy interviewing, in contrast to the standardized interview employed by us (8), Lundy reports the results of a factor analysis of the Interaction Chronograph variables for the interview interactions of his one patient with each of the two therapists. Thus, as was the case in our study, Lundy's use of two therapists with the same *S* enabled him to attempt to cross-validate his findings. One difference between his study and ours was that we used 60 *S*s, each interviewed twice (once by each of two interviewers) while Lundy used one *S* interviewed 19 times by each of two interviewers. In addition, while Lundy's Interaction Chronograph variables differed slightly from those used by us and described above, they were sufficiently similar to permit comparison.

---

[7] These values more correctly should be expressed as proportions rather than as percentages.

Lundy's factor analysis revealed five factors. Three of his factors were identified consistently on replication (using his second set of 19 interviews). Although given slightly different names by Lundy (Speed of Response, Amount of *Therapist* Talk, and Dominance) the first two of these are clearly similar to our Factors I and II; i.e., our patient's "average duration of silence" before he spoke again, and "average duration of *patient* action." Superficially, this latter "duration of patient (talk) action" factor, which we found as our Factor II, may differ from Lundy's own Factor II, "amount of therapist (talk) action." However, since our therapist was instructed always to speak in 5-second utterances, thus making the average amount of therapist talk a constant (5 seconds) from one patient to the next, there was no opportunity for this therapist variable to reflect itself in one of our factors.

There is good evidence (8, p. 357), however, for a high positive correlation between amount of talk produced by each of two conversationalists in free conversation (such as in Lundy's unstandardized psychotherapy interviews). Thus one might hypothesize that had our interviewers been allowed to talk freely, as did Lundy's, and had we then recorded durations of therapist actions, we also might have found a high factor loading for amount of therapist talk on our Factor II. In partial support of this is Lundy's report that his single patient talked more with the therapist who talked more, and talked less with the therapist who talked less. This positive covariation which Lundy reports, while possibly surprising, confirms our own and Goldman-Eisler's similar and independent observation (8, pp. 357–358) and thus adds support to our belief that the reverse is also true; i.e., that a therapist talks more with a patient who talks in long utterances, and less with one who speaks for brief durations. If this is true, Lundy's second factor and ours appear to be identical.

Lundy's third factor, Dominance, differs from our third factor, which we chose to call Pt.'s Adjustment. Since our third factor was the weakest, and since our earlier results have shown that our Dominance variable (and presumably Lundy's also) contains considerable measurement error variance, we are not surprised at the lack of agreement between our own third factor and Lundy's. However, Lundy does report that his Pt.'s Adjustment variable also had a high loading on his third factor, and thus his third factor may be similar to ours.

Lundy's two weakest and also not cross-validated fourth and fifth factors were Frequency of Client Response and Client Variability. The former is similar to our Pt.'s Units variable and thus may be similar to our second factor which, as will be remembered, though called "action" by us was also the factor most highly saturated with our frequency, or Pt.'s Units, variable. Lundy's last factor, Client Variability, is one which is made up of such derived or second-order indices as the variation of the first-order variables Tempo and Activity from one part of the interview to another. Since to date we have not employed such an interview measure we cannot compare our results with Lundy's in regard to this factor.

The results of Lundy's factor analysis and our own are interesting for one additional reason. Although each study was replicated, Lundy used one *S* interviewed 19 times; we used 60 *S*s interviewed once. Interestingly, the results of both studies indicate that duration of silence and duration of action are the most stable (and independent) factors in two-person communicative interactions.

Summary

This study employed 60 outpatient *S*s interviewed under conditions of a previously described standardized interview. The results indicate that as viewed from an Interaction Chronograph framework, doctor-patient inter-

actions (and possibly most other two-person interactions) consist of two very stable factors for any given individual: (*a*) how long on the average he or she waits or remains silent before communicating, and (*b*) the number and average duration of each of these communicative interactions. A third factor, the frequency with which one initiates or starts again with another communication unit of his own when his partner has not answered him, while weak, nevertheless may prove to be more stable as the length of time during which it is assessed (only during Period 2 in our partially standardized interview) is increased. A fourth factor, the efficiency with which a member of the communicating pair synchronizes and adjusts (or maladjusts) to his partner, was evident as a factor (in both our study and Lundy's), but it, too, was not as clear-cut as were the first two factors. The same holds true for a weak and possibly fifth factor, Dominance. Doctor-patient or other two-person patterns of dominance-submission may reflect an additional important interpersonal dimension and thus this fifth factor-variable (again measured in only one subperiod of our present partially standardized interview) will require the development of more refined methods of measurement, or an extension of the total time during which our dominance variable is measured.

Further evidence for the validity, or psychological meaningfulness, of four of these factors (all but Pt.'s Adjustment) comes from a recently completed and cross-validated study of the psychological test correlates of our interview Interaction Chronograph measures (12). In terms of the standard psychological test measures (Rorschach, WAIS, Taylor Anxiety Scale, etc.) used in the latter study, different psychological test-interview correlations were found for patients who: (*a*) were high or low in their interview silences, (*b*) were high or low in their interview actions, (*c*) showed a high or low frequency of initiatives, and (*d*) submitted to or dominated the interviewer.

A second study (13), now undergoing cross-validation, between certain aspects of the verbal content of the interviews and the overt Interaction Chronograph scores, likewise has identified interview content (descriptions of self and others) differences between patients who: (*a*) talk for long or short durations, (*b*) have long silences and corresponding high negative maladjustments, (*c*) show a high or low frequency of initiatives, and (*d*) earn high or low dominance scores.

Thus it would appear that further evidence for the validity of the factors identified by our cross-validated factor analysis is coming from a variety of additional sources. It is our belief that the results of the present study represent merely the smallest number of factors which can account adequately for interview interaction behavior, as this is assessed solely by the Interaction Chronograph. It is expected that a factor analysis which included not only intercorrelations of the 12 Interaction Chronograph variables herein used but also the additional intercorrelations of each of these variables with such other data as psychological test scores, interview content measures, Bales Interaction Process scores, etc. would yield still further stable interview interaction factors, and thus add to our knowledge of interpersonal behavior.

Received July 28, 1958.
Early Publication.

References

1. American Psychological Association. Technical recommendations for psychological tests and diagnostic techniques. *Psychol. Bull.*, 1954, **51** (2, Pt. 2), 1–38.
2. Chapple, E. D. The Interaction Chronograph; its evolution and present application. *Personnel,* 1949, **25**, 295–307.
3. Chapple, E. D. The standard experimental (stress) interview as used in Interaction Chronograph investigations. *Hum. Organization,* 1953, **12** (2), 23–32.
4. Chapple, E. D., Chapple, M. F., & Repp, J. A. Behavioral definitions of personality and temperament characteristics. *Hum. Organization,* 1954, **13**, 34–39.
5. Cronbach, L. J., & Meehl, P. E. Construct validity in psychological tests. *Psychol. Bull.,* 1955, **52**, 281–302.
6. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1954.
7. Lundy, B. W. Temporal factors of interaction in psychotherapy. Unpublished doctoral dissertation, Univer. of Chicago, 1955.
8. Matarazzo, J. D., Saslow, G., & Matarazzo, Ruth G. The Interaction Chronograph as an instrument for objective measurement of interaction patterns during interviews. *J. Psychol.,* 1956, **41**, 347–367.

9. Matarazzo, J. D., Saslow, G., & Guze, S. B. Stability of interaction patterns during interviews: A replication. *J. consult. Psychol.*, 1956, **20**, 267–274.
10. Matarazzo, J. D., Saslow, G., Matarazzo, Ruth, & Phillips, Jeanne S. Stability and modifiability of personality patterns during a standardized interview. In P. A. Hoch & J. Zubin (Eds.), *Psychopathology of communication.* New York: Grune & Stratton, 1958. Pp. 98–125.
11. Matarazzo, J. D., Saslow, G., Matarazzo, Ruth G., & Phillips, Jeanne S. Differences in interview interaction patterns among five diagnostic groups. Paper read at American Psychological Association, New York, N. Y., September, 1957.
12. Matarazzo, Ruth G., Matarazzo, J. D., Saslow, G., & Phillips, Jeanne S. Psychological test and organismic correlates of interview interaction behavior. *J. abnorm. soc. Psychol.*, 1958, **56**, 329–338.
13. Phillips, Jeanne S. The relationship between two measures of interview behavior comparing verbal content and verbal temporal patterns of interaction. Unpublished doctoral dissertation, Washington Univer. (Missouri), 1957.
14. Phillips, Jeanne S., Matarazzo, J. D., Matarazzo, Ruth G., & Saslow, G. Observer reliability of interaction patterns during interviews. *J. consult. Psychol.*, 1957, **21**, 269–275.
15. Saslow, G., Matarazzo, J. D., & Guze, S. B. The stability of Interaction Chronograph patterns in psychiatric interviews. *J. consult. Psychol.*, 1955, **19**, 417–430.
16. Saslow, G., Matarazzo, J. D., Phillips, Jeanne S., & Matarazzo, Ruth G. Test-retest stability of interaction patterns during interviews conducted one week apart. *J. abnorm. soc. Psychol.*, 1957, **54**, 295–302.
17. Saslow, G., & Matarazzo, J. D. A technique for studying changes in interview behavior. In E. A. Rubinstein & M. B. Parloff (Eds.), *Research in psychotherapy*. Washington, D. C.: Amer. Psychol. Ass., 1959, in press.
18. Thurstone, L. L. *Multiple factor analysis.* Chicago: Univer. Chicago Press, 1947.

Ednl. Research
Dated
Acc No

Journal of Consulting Psychology
Vol. 22, No. 6, 1958

Problems of Reliability in Observing and Coding Social Interactions[1,2]

Allen T. Dittmann
National Institute of Mental Health

In observing social interactions and coding reports of the observations, problems of reliability arise at every turn. This paper reports a series of reliability studies which have been carried out at the Child Research Branch, NIMH, where a variety of children have been observed in "naturalistic" settings within a residential treatment framework. Various aspects of the findings are relevant to different substantive studies within the research program, but the reliability tests are better understood when presented as a whole.

The basic data are reports of 5 to 10 minute observations of children in social situations, dictated from memory immediately after the observations. The protocols are coded in the "Circle" of interpersonal mechanisms (1) by pairs of coders working together. Reliability of coding in this paper means agreement between two independent coder pairs, or one pair repeating coding with at least six months intervening. Reliability of observing means two observers reporting independently on the same situation.

In judging which behaviors reported by the observer are interpersonal rather than self-directed acts, agreement is moderately high. Disagreements occur where the observer's descriptions lack detail and where the interactions involved in the acts are equivocal.

The reliability of coding protocols item by item is far greater than chance expectancy, and yet the degree of association is only moderately high. The reliability of profiles made up of the codings of many interactions, on the other hand, is very high, both in repeat reliability and in agreement of independent pairs of coders. The intensity dimension yields similar results: agreement is clearly beyond chance, but only groups of many such ratings can be used with confidence.

Reports of two observers reporting on the same situations have also been studied. While there is only general agreement on individual acts and on the sequence of events, profiles of codings based on the reports are just as similar as are those based on independent codings of the same reports.

Conclusions: The reliability of these measures is analogous to that of tests. Items of low reliability may be combined to yield tests of high reliability. Where profiles of individuals or groups are to be compared, the system we have used is very adequate. Where sequences of individual acts are to be studied, however, as in the details of symptomatic behavior, the number of sequences must be large enough that the low item reliability does not detract from any generalizations drawn from the findings.

Brief Report.
Received June 13, 1958.

Reference

1. Freedman, M., Leary, T., Ossorio, A., & Coffey, H. The interpersonal dimension of personality. *J. Pers.*, 1951, **20**, 143–161.

---

[1] An extended report of this study may be obtained without charge from Allen T. Dittmann, National Institute of Mental Health, Bethesda 14, Maryland, or for a fee from the American Documentation Institute. Order Document No. 5740, remitting $1.25 for microfilm or $1.25 for photocopies.

[2] The author wishes to thank D. Wells Goodrich and Harold L. Raush for their help throughout these studies.

Journal of Consulting Psychology
Vol. 22, No. 6, 1958

Environmental Factors and Outpatient Clinic Intake

Marvin Hyman and Julian Wohl
VA Mental Hygiene Clinic, Detroit

In the everyday functioning of a mental hygiene clinic or agency, marked fluctuations can be observed in the number of people who present themselves for the available services. Each agency develops its own explanation for this phenomenon with the result that there currently exists a number of untested hypotheses to explain variations in daily intake. These hypotheses are varied and sometimes colorful. The authors have heard explanations that hypothesize relationships between daily intake and weather conditions, unemployment, atom bomb explosions, crime rate, sputniks, etc. Although these hypotheses have been untested in the main, they are nevertheless the product of observations of experienced clinical workers and deserve some evaluation.

It was the purpose of this study to examine systematically the daily and monthly variations in the number of persons who present themselves for mental hygiene services and to investigate the possible relationships that existed between these variations and (*a*) certain climatological conditions and (*b*) certain aspects of the employment situation. Specifically, the following hypotheses were tested: (*a*) daily intake varies systematically according to the day of the week and the month of the year, (*b*) daily intake is related to daily climatological conditions, e.g., temperature, precipitation, sunshine, etc., (*c*) daily intake is correlated with the rate of local unemployment.

In order to test the foregoing hypotheses, the daily intake of the Detroit VA Mental Hygiene Clinic for the year 1956 was studied. Daily intake was defined as the number of individuals presenting themselves on a particular day, without appointment, for the various services offered by the clinic. Patients who were seen on the basis of regularly scheduled appointments were not included in the count.

From information supplied by the U. S. Department of Commerce Weather Bureau (1), it was possible to select three variables which grossly reflected weather conditions on a particular day. These were: Average temperature, total precipitation, and percentage of possible sunshine.

The Michigan Employment Security Commission[1] supplied two indices of the employment situation in the Detroit Metropolitan Area for 1956: Total unemployment estimate (by months) and the number of claims filed for unemployment compensation (by weeks).

Results

In 1956 there were 252 working days. The median intake over the year was 7.43 patients per day. Reference to Table 1 reveals that heavier intake (more than seven patients) was most likely to occur on Mondays, and least likely on Wednesdays. Similarly, reference to Table 2 indicates that heavier intake days were most likely to occur in the months of September, November, and March; and

Table 1

Chi Square Test of Intake Variation by Day of the Week

	Day of the week				
Intake	M	T	W	Th	F
Under the median	15	20	37	28	28
Over the median	34	31	12	23	24

Note.—Chi square = 23.238; $df = 4$; $p = .001$.

---

[1] Personal communication.

Table 2

Chi Square Test of Intake Variation by Month of the Year

Month	Number of working days intake was—	
	Under median	Over median
Jan.	12	9
Feb.	13	7
Mar.	8	14
Apr.	8	13
May	11	11
June	15	6
July	17	4
Aug.	14	9
Sept.	3	16
Oct.	13	10
Nov.	6	14
Dec.	11	8

Note.—Chi square = 30.281; $df = 11$; $p = .003$.

were least likely to occur in the months of June, February, and December.

The effect of weather conditions upon intake was submitted to chi square analyses with the following results: Temperature was related to intake in a curvilinear fashion, i.e., heavier intake tended to occur on those days of the year when the temperature was moderate (second and third quartiles of the temperature distribution) and was least likely when it was very cold or very warm (first and fourth quartiles). The chi square was 4.036 with 1 df, $p = .04$. The presence or absence of rain, snow, and/or sunshine seemed not to be significantly related to intake.

The analysis of the relationship between intake and conditions of employment revealed that intake is unrelated either to the number of unemployed during the year (rho = −.04), or to the number of claims for unemployment compensation that are filed (rho = .02).

We suspect, but have no data to support this, that Monday is the heaviest intake day because of a backlog of applicants which accumulates over the weekend. Another guess is that Wednesday typically has few intakes because of a phenomenon peculiar to Detroit. On that day in Detroit most physicians do not maintain office hours and private clinics are closed. We would guess that the community has generalized from this fact and lives as if it applied to the VA Clinic.

The mythology of the outpatient clinic has suggested that unemployment in an industrial, one-industry area would affect intake activity in an outpatient mental health agency. The agency, so runs this argument, deals primarily with marginal people of low socioeconomic class who would be more likely to come to the agency when times get hard. Although it is true that the clinic population is comprised largely of individuals from low socioeconomic classes, the results of this study show that whatever accounts for fluctuation in requests for clinic services, it is not the labor situation. Like the U.S. mail, our clientele was not discouraged by rain, sleet, snow, or hail, although when the weather became very hot, it tended, unlike Noel Coward's Englishmen, "to stay out of the midday sun," and when it was extremely cold, wisdom overcame anxiety, and the clients remained close by what passed for their firesides.

From these results, it can be concluded that, by and large, the applicants for the services of this clinic come because of needs for help which are not based on situational factors such as unemployment and weather conditions.

Summary

This study tested the hypotheses that weather and employment conditions affected the applications for treatment in a VA mental hygiene clinic. With some exceptions, the obtained data did not support these hypotheses.

Received November 15, 1957.

Reference

1. U. S. Department of Commerce, Weather Bureau. *Local climatological data (monthly reports).* Washington, D. C.: U. S. Government Printing Office, 1956.

Journal of Consulting Psychology
Vol. 22, No. 6, 1958

A Factor Analysis of Indices from Four Projective Techniques Associated with Four Different Types of Physical Pathology

Claire M. Vernier
Veterans Administration Hospital, Baltimore, Maryland

John W. Stafford
Catholic University of America

and Arnold D. Krugman
Veterans Administration Center, Martinsburg, West Virginia

The analysis to be reported represents a partial approach to two current clinical problems: (*a*) statistical validation of relationships among specific projective test scores or signs; and (*b*) measurement of relationships between psychological variables and different types of medically diagnosed physical pathologies.

Both of these somewhat more theoretical questions were derived from an immediate practical problem of the need for a set of rigidly definable and defensible bases for the identification of specific test variables to be used in a comparison of significant psychological differences between two groups of equally physically handicapped male veterans (3).

While several factorial studies of projective techniques have been reported in the literature, these analyses have been directed primarily to the structure of a given test rather than to the intertest factorial composition of individual psychological variables. Thus, there are Guertin's analyses of the Szondi and Bender-Gestalt (4, 5, 6, 7, 8, 9) and the studies of the Rorschach (1, 2, 10, 11), but as far as the authors know there have been no factor analysis studies reported of variables from more than one projective technique.

The literature describing relationships between physical pathology and personality functioning is vast. Extensive efforts have been made to demonstrate personality types associated with particular physical disabilities. In less intensive fashion, a great deal of work has occurred describing the relationship between kind of disability and adjustment to it. Numerous papers have appeared regarding prediction of specific behaviors in the areas of hospital adjustment, response to medical treatment, or posthospital adjustment. However, in none of these reports were medical and psychological data integrated into a correlational matrix and factored.

Therefore, the major purpose of this study was to verify by means of multiple factor analysis, the relationships among the projective test scores and signs previously considered to be measures of seven specific psychological variables. A secondary purpose was to determine, also by means of multiple factor analysis, the relationships among four types of physical pathology and these same projective test scores.

Procedure

The total sample consisted of 105 male members of the VA Center, Martinsburg, West Virginia Domiciliary and 45 equally physically disabled male veterans residing in West Virginia who had never applied for admission to any VA Domiciliary or similar type

Table 1

Variables Included in Initial Correlation Matrix

Variable	Test	Definition
A. Test Scores from the four projective tests:		
Withdrawal		
1[a]	Bender-Gestalt	Upper ⅓ or left margin placement of all 9 figures
2	Bender-Gestalt	Figure #2, the line of 3 circles straight or left slanted, not right
3	Person Drawing	Hands omitted
4[a]	Rorschach	Total % human content
5	Rorschach	Total sum color ($\frac{1}{2}$FC+1CF+1.5C)
6	Sentence Completion	Total press-rejection score
7	Sentence Completion	Total need-affiliation score
8[a]	Sentence Completion	Total need-infavoidance score
Energy or Activity		
9	Person Drawing	Legs or arms uneven in length or width
10	Rorschach	Total animal movement score (FM)
11[a]	Rorschach	Total sum of large details (D)
12[a]	Bender-Gestalt	Figure #6, 1½ times or more enlarged
13[a]	Bender-Gestalt	Figure #8, 1½ times or more enlarged
Feelings of Invalidism		
14[a]	Rorschach	Total animal object content score
15	Rorschach	Total anatomical content score
16	Rorschach	Total plant content score
17	Rorschach	Total number of dysphoric responses[b]
18	Person Drawing	Unseeing eye (drawn as empty circle)
19	Person Drawing	Body shading other than hair or clothing details present
20	Person Drawing	Legs or arms drawn as single straight line[c]
21[a]	Person Drawing	Legs omitted
22	Sentence Completion	Press-physical lack
Anxiety (Tension)		
23	Rorschach	Total number of cards rejected
24	Rorschach	Total score of inanimate movement (m)
25	Rorschach	Total score of diffuse shading (K)
26[a]	Rorschach	Total score of shading as two-dimensional perspective (k)
27	Bender-Gestalt	Erasures present in one or more figures
28	Bender-Gestalt	Retracing present in one or more figures
29[a]	Person Drawing	Person drawn as off-balance
30	Sentence Completion	Total score need-defendance
Depression		
31	Rorschach	Sum of achromatic color (C′)
32[a]	Bender-Gestalt	Figure #2, horizontal slanted down to right
33	Bender-Gestalt	Figure #3 rotated down (45° to 225°)
34[a]	Bender-Gestalt	Figure #6, horizontal curve slanted down to right
Dependency–Passivity		
35	Rorschach	Total sum of oral content
36[a]	Bender-Gestalt	Figure #7, left side rotated down
37[a]	Person Drawing	Center line of buttons indicated
38[a]	Sentence Completion	Total score press-nurturance
39	Sentence Completion	Total score need-passivity
40[a]	Sentence Completion	Total score need-succorance
Ambition		
41	Rorschach	Total sum of small details (d)
42[a]	Rorschach	Total number of responses (R)
43	Rorschach	Eagle or emblem content present in one or more cards
44	Person Drawing	Additional non-clothing details present
45	Sentence Completion	Total score need-achievement
46	Sentence Completion	Total score need-recognition
47[a]	Person Drawing	Neck omitted
B. Other Variables:		
48	Orthopedic pathology present	
49	Respiratory pathology present	
50	Cardiac pathology present	
51	Neurological pathology present	
52	IQ score from group-administered Otis S.A., Intermediate level, Form A	
53	Physical fitness test score	

[a] Variables subsequently dropped from the factor analysis matrix.

[b] As specifically defined concepts of physical destruction or decay, such as germs, rotten tree, corroded by rust, eaten away tissue, etc.

[c] Except when part of a total stick (line) figure.

of institutional living.[1] Each subject had a definite medical diagnosis of one or more of the following four disease processes: orthopedic, as verified by X-rays; respiratory, as verified by X-rays and other laboratory studies; cardiac, as verified by both electrocardiograph and blood studies; neurological, as verified by either spinal tap or electroencephalogram. Incidence of pathology for the total sample of 150 cases was: orthopedic, 43%; respiratory, 28%; cardiac, 38%; neurological, 16%.

A completed battery of the following five psychological tests, group administered, was available for each of the 150 cases: Otis S.A. Test of Mental Ability, Intermediate Examination, Form A; Bender-Gestalt figure drawings (from slides); Rorschach; Drawing of a Person; and the Rhode-Hildreth Sentence Completion Test. One other measure also available was a quantitative score of physical fitness capacity derived from a performance test developed by the Chief of Physical Medicine and Rehabilitation (3).

From the four projective tests 47 scores were selected which, on the basis of previous studies and of clinical experience, presumably reflect personality variables of significance in relation to the comparison of differences between the Domiciliary members and the community control sample. Such variables include: withdrawal, energy, feelings of invalidism, anxiety, depression, dependency-passivity, and ambition. Table 1 summarizes the 53 variables in the original correlation matrix.

Tetrachoric correlations were computed among all of the 53 variables. In the case of quantitative test scores, the median was used as the cutting point. In order to have a matrix of a size more feasible to factor, 18 of the 47 psychological test scores were dropped either because of essentially zero correlations or correlation coefficients above .75, with one or more of the variables included.

The 35 × 35 matrix was factored by the Thurstone Complete Centroid Method. The factor loadings, after ten oblique rotations, are given in the following section. A cutting point of ±.35 has been used, except in three cases where the pattern of correlations seemed to justify going down to .32 or .33. Listed first are the variables loading uniquely on the given factor; then are listed, with the variable number in parentheses, those variables loading ± .35 or above on other factors also.

Results

Factor A

No.	Variable	Loading
3	Hands omitted (Person Drawing)	.63
7	Need-affiliation (Sentence Completion)	−.49
[23]	Card rejections (Rorschach)	−.64
[17]	Dysphoric responses (Rorschach)	.53
[20]	Single line legs or arms (Person Drawing)	−.46
[51]	Neurological pathology present	.36

On the basis of the two test variables which have significant loadings for this one factor alone, Factor A is tentatively identified as a "withdrawal" variable. The significant positive loading for neurological pathology present is consistent with clinical observations.

Factor B

No.	Variable	Loading
27	Erasures (Bender-Gestalt)	.49
22	Press-physical lack (Sentence Completion)	−.47
[51]	Neurological pathology present	.72
[30]	Need-defendance (Sentence Completion)	.68
[9]	Uneven legs or arms (Person Drawing)	.55
[15]	Anatomical content (Rorschach)	−.54
[24]	High inanimate movement (Rorschach)	.54

Factor B is tentatively identified as a "tension" factor. The negative loadings for two scores usually interpreted as methods of defense against anxiety [22, 15] would be consistent with this hypothesis.

Factor C

No.	Variable	Loading
53	Physical fitness test score	.60
44	Additional details (Person Drawing)	.41
41	Small details (Rorschach)	.37
6	Press-rejection (Sentence Completion)	.33
[43]	Eagle or emblem content (Rorschach)	.59
[46]	Need-recognition (Sentence Completion)	.56
[20]	Single line legs or arms (Person Drawing)	−.47
[23]	Card rejections (Rorschach)	−.47
[9]	Uneven legs or arms (Person Drawing)	−.41
[35]	Oral content (Rorschach)	.41

The combination of variables with significant loadings on Factor C would suggest a tentative identification as "needs for social approval or attention" rather than either energy or ambition.

---

[1] The community control group was obtained through the cooperation of the West Virginia State Board of Vocational Education, Division of Vocational Rehabilitation.

Factor D

No.	Variable	Loading
25	Diffuse shading (Rorschach)	.32
[19]	Body shading (Person Drawing)	.76
[28]	Retracing (Bender-Gestalt)	.68
[15]	Anatomical content (Rorschach)	−.56
[49]	Respiratory pathology present	−.50

Factor D would appear to be an "anxiety" factor as differentiated from "physical tension" reflected in Factor B.

Factor E

No.	Variable	Loading
39	Need-passivity (Sentence Completion)	.39
[51]	Neurological pathology present	.93
[28]	Retracing (Bender-Gestalt)	.60
[30]	Need-defendance (Sentence Completion)	.60
[19]	Body shading (Person Drawing)	.41
[43]	Eagle or emblem content (Rorschach)	.41
[9]	Uneven legs or arms (Person Drawing)	.39
[17]	Dysphoric responses (Rorschach)	.38

On the basis of the variable with the highest unique loading for Factor E, it is tentatively identified as "dependency-passivity."

Factor F

No.	Variable	Loading
5	High weighted sum color (Rorschach)	.71
16	Plant content (Rorschach)	.42
[9]	Uneven legs or arms (Person Drawing)	−.46
[49]	Respiratory pathology present	−.34

Factor F is possibly an indication of affective lability.

Factor G

No.	Variable	Loading
52	IQ	.73
2	Figure #2, straight or slanted to left (Bender-Gestalt)	.45
50	Cardiac pathology present	.44
10	High animal movement (Rorschach)	.41
[19]	Body shading (Person Drawing)	.65
[24]	High inanimate movement (Rorschach)	.43

The combination of variables with significant loadings on this factor is clinically interesting as well as puzzling to some extent. One hypothesis is that there may be a kind of system of pseudo-intellectual defenses available to persons with better than average intelligence.

There were five variables, 18, 31, 33, 45, and 48, which had no significant loading on any of the seven factors.

The seven factors were intercorrelated. Most of the correlations were low and negative. Only two, between B and E (−.49) and between C and F (−.44), were above .35. It is, therefore, quite unlikely that any second-order factors would be found.

Discussion of Results

The factor structure that emerges from the oblique rotations appears to be fairly well defined. Of the originally postulated seven psychological variables for which projective test scores were selected, three were relatively easily identifiable as factors, namely: withdrawal (Factor A), anxiety (Factor D), and dependency-passivity (Factor E). Four of the postulated psychological variables did not emerge as factors, depression, ambition, energy, and feelings of invalidism. While it is possible that adequate projective test indices of these variables were not included in the factor matrix, it is of interest that three somewhat related factors were identified: affective lability (Factor F), needs for social approval or attention or prestige (Factor C), and pseudo-intellectual defenses (Factor G). Not previously anticipated was the appearance of a physical tension factor (Factor B) as differentiated from an anxiety factor.

It can be noted that neither a test nor a medical pathology type of factor structure was found. That is, there was not found a series of factors each loaded with indices from a single test, or four factors each with one medical pathology variable showing a high loading. The pattern of factor loadings for the four types of medical pathology would be consistent with the theory that personality variables are not associated with specific types of physical disease. It may very well be that the personality variables are related to kinds of reaction to illness or ability, although this can not be concluded from evidence in the present study.

We find that orthopedic pathology has no significant loadings on any of the seven factors; respiratory pathology has negative loadings on Factors F and D; cardiac pathology shows a positive loading on Factor G; neurological pathology has high loadings on Factors A, B, and E.

Summary

Forty-seven scores or indices were selected from four different projective tests and intercorrelated; also included in the correlation matrix were medical diagnoses of four types of physical pathology, IQ score, and a score of physical fitness capacity. Eighteen of the test scores were eliminated, either because of zero correlations or very high correlations with other variables, and the remaining 35 × 35 matrix factored.

Seven factors emerged which are tentatively identified as: withdrawal, physical tension, needs for social approval or prestige, anxiety, dependency-passivity, affective lability, and pseudo-intellectual defenses. The factor structure seemed unrelated to either the kind of test or the type of medical pathology.

Presence of orthopedic pathology did not show a significant loading on any of the seven factors. Respiratory pathology present had significant loadings on not labile and not anxious. Cardiac pathology present showed a significant loading on intellectual defenses; neurological pathology present had high loadings on withdrawal, physical tension, and dependency-passivity.

The results of the analysis would tend to indicate a partial validation for some of the clinical interpretations made from selected projective test variables and to support the hypothesis that personality factors are not associated with specific types of physical disease.

Received June 9, 1958.
Early Publication.

References

1. Adcock, C. J. A factorial approach to Rorschach interpretation. *J. gen. Psychol.*, 1951, **44**, 261–272.
2. Cox, S. M. A factorial study of the Rorschach responses of normal and maladjusted boys. *J. genet. Psychol.*, 1951, **79**, 95–115.
3. Fogel, E. J., Swepston, E. R., Zintek, S. S., Vernier, Claire M., Fitzgerald, J. F., Marnocha, R. S., & Weschler, C. H. Problems of the aging: Conclusions derived from two years of interdisciplinary study of domiciliary members in a veterans administration center. *Amer. J. Psychiat.*, 1956, **112**, 724–730.
4. Guertin, W. H. A factor analysis of some Szondi pictures. *J. clin. Psychol.*, 1951, **7**, 232–235.
5. Guertin, W. H. A factor analysis of the Bender-Gestalt tests of mental patients. *J. clin. Psychol.*, 1952, **8**, 362–367.
6. Guertin, W. H. A factor analysis of curvilinear distortions on the Bender-Gestalt. *J. clin. Psychol.*, 1954, **10**, 12–17.
7. Guertin, W. H. A transposed analysis of the Bender-Gestalts of brain disease cases. *J. clin. Psychol.*, 1954, **10**, 366–369.
8. Guertin, W. H. A transposed factor analysis of schizophrenic performance on the Bender-Gestalt. *J. clin. Psychol.*, 1954, **10**, 225–228.
9. Guertin, W. H. A transposed analysis of the Bender-Gestalts of paranoid schizophrenics. *J. clin. Psychol.*, 1955, **11**, 73–76.
10. Williams H. L., & Lawrence, J. F. Comparison of the Rorschach and MMPI by means of factor analysis. *J. consult. Psychol.*, 1954, **18**, 193–197.
11. Wittenborn, J. R. A factor analysis of discrete responses to the Rorschach Ink Blots. *J. consult. Psychol.*, 1949, **13**, 335–340.

Journal of Consulting Psychology
Vol. 22, No. 6, 1958

Rorschach Maturity Scores of the Mothers of Schizophrenics[1]

C. L. Winder
Stanford University

and Robert E. Kantor
Palo Alto, California

A number of studies have pointed to the impact of the "schizophrenogenic" parents, especially mothers (e.g., 3, 5, 8, 12, 16, 17, 18, 19), and other studies are indicative of pathological configurations in Rorschach results of relatives of schizophrenics (e.g., 13, 14). Further examination of the proposition that pathology can be discovered in the Rorschach performances of mothers of schizophrenics is in order. Such mothers are expected to show relatively severe manifestations as compared with mothers of nonschizophrenic offspring.

In 1950, Prout and White published one of the few studies in which there was a comparison of reasonably well matched groups of mothers of schizophrenics and mothers of nonschizophrenics. Prout and White studied, by means of personal interview, the dominant orientations of all these mothers toward their sons in such areas as child-rearing attitudes and also recall of the various actual events and adjustments of the growing child. Additionally, a Rorschach test was administered to each mother in an effort to learn something about her personality pattern. Essentially, the findings of Prout and White were these: (*a*) the mothers of the schizophrenics seemed to be overprotective and overcontrolling in contrast to the control mothers, (*b*) the experimental mothers had somewhat higher and more unrealistic levels of aspiration for their sons and subordinated the happiness of the child to such considerations, (*c*) yet, paradoxically, their demands on their sons were somewhat less explicit than the demands made by the control mothers, and (*d*) clinical assessments of the Rorschachs indicated that the experimental mothers manifested more pathological personality configurations.

It is this last finding which leads to the effort to make a more particularistic statement of the difference in personality characteristics, based on the Rorschach performances of these two groups of mothers. The Prout and White data can be utilized in a test of the hypothesis that the "level of perceptual organization" aspect of the complex pathology called schizophrenia is more evident in the mothers of schizophrenics than among mothers of normals.

Method

The Prout and White sample consisted of 50 women, half of them being the experimental group who were mothers of young adult males diagnosed as schizophrenic, and the other half consisted of a control group of mothers of normal sons. In general, Prout and White made the two groups comparable on several usual matching variables. However, there were some differences in chronological age and about one-third of the experimental mothers had been born in foreign countries. The Rorschach protocols were coded by White and sent to the present authors who rated each protocol for "personality-age-level."

---

[1] We express our gratitude to C. T. Prout and Mary A. White, of the New York Hospital, Westchester Division, White Plains, New York. Their data and cooperation made the study possible.

"Personality-age-level" ratings are based on the viewpoint that the Rorschach performance can be scored for maturity of personality development by rating the protocol on a scale (a five-point scale in the present study) on which points are defined by the average perceptual organization levels of successive representative chronological age groups beginning with very young children and extending to the adult group. In other words, it is assumed that the level of organization of Rorschach responses is associated with the level of organization of adjustment in life. Since the rating points are defined in terms of average Rorschach performances, the definitions of rating points are limited by the adequacy of normative data. In the present study, norms relied upon included those of Halpern (6), Kay and Vorhaus (9), Klopfer and Margulies (10), McFate and Orr (11), and Cass and McReynolds (2).

After each judge had rated every protocol (there were 25 experimental and 23 control records available), the key to the code was obtained and the statistical tests for differences between the two groups of mothers were made.

Results

The judges were permitted any gradations between 1 and 5, or in other words were not required to use only the points 1, 2, 3, 4, or 5. The product-moment correlation between the two sets of judges' ratings is .83. A mean rating based on the two judges' ratings for each protocol was adopted as the best available score. On the five-point scale where 5 is the most immature and 1 is the most mature, the mean rating for the experimental group is 2.89 and that for normals is 2.43. The difference between means yielded a t of 3.08, $p < .01$. There is considerable overlap of the groups, as is expected. Taking the median of the combined groups as a reference point, 10 mothers of schizophrenics and 14 mothers of normals are above, and 15 mothers of schizophrenics and 9 mothers of normals are below, the median.

For each of the two sets of ratings, a chi-square test for relationship between group and rating was made. The four rating categories of the 2 × 4 chi-square table were the quartiles of the combined distribution of the 48 ratings made by each judge. For one judge, the chi square was significant between the .02 and .01 levels. For the other judge, the significance level was between .05 and .02.

Discussion

Since there is a statistically significant difference between the mean age-level ratings for the experimental and control groups, it is concluded that among mothers of schizophrenic sons there is a disproportionate number who manifest pathology. This pathology is tentatively defined as immaturity of perceptual organization. It is assumed that this immaturity is associated with detrimental developmental conditions and is manifest in pathogenic conduct in parent-child relationships which has contributed significantly to the schizophrenic reactions of the sons.

Since the distributions of the highly reliable ratings show considerable overlap, it is assumed that the experimental mothers are only quantitatively deviant from the culturally normal range of conduct. Several mothers of nonschizophrenic sons manifest severe pathology. Their less debilitating effect is explained on the basis of assumed compensatory relationships of their sons with others. As some experimental mothers manifest minimal pathology, alternative hypotheses must be invoked to account for the development of schizophrenic reactions in these sons.

The approach used in evaluating the Rorschachs is apparently very similar to a system developed at Clark University and reported by Friedman (4), Siegel (18), Wyatt (20), Pena (15), and Hemmendinger (7). A similar system was used by Becker (1). Since the direct comparison of scoring one set of protocols by all of these methods was not carried out, no firm conclusion is justified, as yet, regarding the comparability of such scores.

Summary

Rorschach protocols of 48 women, 25 of whom were mothers of young adult male schizophrenics and 23 mothers of normal

sons, were scored for personality-age-level, i.e., maturity of personality.

Within the limitations of this study, the following conclusions are drawn:

1. Personality-age-level Rorschach scoring can be done with good reliability.

2. As a group, mothers of schizophrenic sons manifest more pathology than a group of mothers of normal sons.

Received December 2, 1957.

References

1. Becker, W. C. A genetic approach to the interpretation and evaluation of the process-reactive distinction in schizophrenia. *J. abnorm. soc. Psychol.,* 1956, **53**, 229–236.
2. Cass, W. A., & McReynolds, P. R. A contribution to Rorschach norms. *J. consult. Psychol.,* 1951, **15**, 178–184.
3. Freeman, R. V., & Grayson, H. M. Maternal attitudes in schizophrenia. *J. abnorm. soc. Psychol.,* 1955, **50**, 45–52.
4. Friedman, H. Perceptual regression in schizophrenia: An hypothesis suggested by the use of the Rorschach test. *J. proj. Tech.,* 1953, **17**, 171–185.
5. Gerard, D. L., & Siegel, J. The family background of schizophrenia. *Psychiat. Quart.,* 1950, **24**, 47–73.
6. Halpern, Florence. *A clinical approach to children's Rorschachs.* New York: Grune & Stratton, 1953.
7. Hemmendinger, L. Perceptual organization and development as reflected in the structure of Rorschach test responses. *J. proj. Tech.,* 1953, **17**, 162–170.
8. Kasanin, J., Knight, E., & Sage, P. The parent-child relationships in schizophrenia. *J. nerv. ment. Dis.,* 1934, **79**, 249–263.
9. Kay, Lillian W., & Vorhaus, Pauline G. Rorschach reactions in early childhood. Part II. Intellectual aspects of personality development. *Rorschach Res. Exch.,* 1943, **7**, 71–77.
10. Klopfer, B., & Margulies, Helen. Rorschach reactions in early childhood. *Rorschach Res. Exch.,* 1941, **5**, 1–23.
11. McFate, Margarite Q., & Orr, Frances. Through adolescence with the Rorschach. *Rorschach Res. Exch.,* 1949, **13**, 302–319.
12. Mark, J. C. The attitudes of the mothers of male schizophrenics toward child behavior. *J. abnorm. soc. Psychol.,* 1953, **48**, 185–189.
13. Modonesi, Carlo. Contributo all'ereditarietà della schizofrenia; studio psycologico su parenti di schizofrenici condotto con l'aiuto del test di Rorschach. *Rass. Studi psichiat.,* 1951, **40**, 118–146. (*Psychol. Abstr.,* **26**: 4141.)
14. Müller, U. G. Gesunde Familien Schizophrener in Rorschach Versuch. *Nervenarzt,* 1950, **21**, 29–35. (*Psychol. Abstr.,* **26**: 5757.)
15. Pena, C. D. A genetic evaluation of perceptual structurization in cerebral pathology: An investigation by means of the Rorschach test. *J. proj. Tech.,* 1953, **17**, 186–199.
16. Prout, C. T., & White, Mary A. A controlled study of personality relationships in mothers of schizophrenic male patients. *Amer. J. Psychiat.,* 1950, **107**, 251–256.
17. Reichard, Suzanne, & Tillman, C. Patterns of parent-child relationships in schizophrenia. *Psychiatry,* 1950, **13**, 189–204.
18. Siegel, E. L. Genetic parallels of perceptual structuralization in paranoid schizophrenia: An analysis by means of the Rorschach technique. *J. proj. Tech.,* 1953, **17**, 151–161.
19. Tietze, Trude. A study of mothers of schizophrenic patients. *Psychiatry,* 1949, **12**, 55–65.
20. Wyatt, F. Some remarks on the place of cognition in ego psychology. *J. proj. Tech.,* 1953, **17**, 144–150.

Journal of Consulting Psychology
Vol. 22, No. 6, 1958

Rorschach Card Rejection by Psychiatric Patients[1]

Arthur S. Tamkin[2]

Veterans Administration Hospital, Northampton, Massachusetts

There is a great deal of inconsistency in the Rorschach literature among statements about the incidence of Rorschach card rejection within the various nosological groups and about the determining factors proposed to account for it. In Rorschach's (6) discussion of the incidence of rejections he stated that normal subjects almost never fail to give a response, while neurotics occasionally reject, and schizophrenics frequently reject the cards. He attributed the neurotic's failure to respond to "inhibitions due to complexes." Beck (1) considered rejections to result from "fearful needs for certainty," and he regarded them as an indication of defensive behavior, the subject thereby trying to shut off either affect or intellect. In the schizophrenic, however, he considered them as symptomatic of regression.

Klopfer and Kelley (3) observed that there were differential rejection rates associated with the various cards. They stated that Card IX was the only one frequently rejected by normal subjects, with Cards II and VI next in order of frequency. It was their belief that the rejection of Cards II, IV, VI, and IX has less pathological significance than rejection of the other cards, that neurotic subjects usually reject these cards, and that the rejection of more than four cards points toward a psychotic disorder. They proposed that while the general significance of rejection is a blocking or resistance against the situation, it may spring from a variety of causes such as an intellectual or emotional incapacity to cope with the situation or an aggressive negativism.

Using the group Rorschach method with psychiatric patients to explore favorable prognostic indications, Scherer (7) found that Rorschach card rejection was inversely related to spontaneous remission, and that there were trends for patients who showed the least number of rejections to be the ones most likely to improve from electroconvulsive therapy and lobotomy. He also found that two weeks after lobotomy, rejections decreased in both individual and group methods of administration (8). In the context of these studies, rejection was considered to signify withdrawal from the immediate environmental situation and a decrease in the strength of ego boundaries.

It is understandable that inconsistent opinions might arise because of the absence of relevant statistical data. However, recent studies reporting empirical data on rejection by psychiatric groups show considerable agreement and elucidate the actual incidence of rejections. In Mensh and Matarazzo's study (4) of 201 hospital patients, it was found that 19% of their psychotics, 24% of the neurotics, and 37% of the organics rejected one or more cards. These values were not significantly different from each other according to the chi square test. Rejections occurred primarily in response to Cards IX, VI, VII, X, and IV, which provoked 83% of all rejections. In a cross-validation study, Sisson, Taulbee, and Gaston (9) found that within their subject groups, rejections were elicited from 33.3% of the neurotics, 32.0% of the schizophrenics, 17.4% of the normal adults, 10.0% of the normal children, and 22.7% of the mentally defective children. These differences were significant for comparisons within the adult groups and the juvenile groups, with the exception that the percentages of rejections by the neurotics and the schizophrenics were not significantly different from each other. When all groups were matched for age and number of responses, the only significant difference with respect to the rejection data was the more frequent rejection of Card X by the schizophrenics than by the normals. They concluded that although Rorschach card rejection has little or no diagnostic significance in terms of the common nosological categories, it might be related to personality variables which cut across the broad diagnostic classifications. It was to explore this possibility that the present study was designed.

---

[1] From the VA Hospital, Northampton, Mass. The author is appreciative of the help in the preparation of this report given by the staff of the Clinical Psychology Service and by Morton Wiener of Clark University and Maria Rickers-Ovsiankina of the University of Connecticut.

[2] Now at the Columbus Receiving Hospital for Children, Columbus 19, Ohio.

Method

Rorschachs of 109 psychiatric patients at the Northampton Veterans Administration Hospital for whom concurrent MMPI data were also available were randomly selected from the files for use in this study. Rejection was defined as the failure to offer a scorable response to any Rorschach card during the performance proper, and the specific cards eliciting a rejection were recorded for each subject. The personality variables investigated as determinants of rejection were defensiveness, psychasthenia, depression, pathology, and psychiatric diagnosis. The first three traits were suggested as promising variables in view of their advocacy by other investigators as determinants of item omission on the MMPI (10), a type of test behavior presumed to be the "objective" test counterpart of Rorschach card rejection. These traits were represented by the following scales of the MMPI: "Cannot Say," *K*, *L*, *Pt*, and *D*. The *F* scale was taken as the measure of pathology, and psychiatric diagnosis was either neurosis or psychosis (predominantly schizophrenia) previously established by a diagnostic conference for a subgroup of 96 patients. All MMPI scales were dichotomized at their respective medians, and Rorschach rejection was dichotomized on the basis of its absence or presence. A series of two by two contingency tables was drawn up, and chi square tests were computed separately for each measure.

Table 1

Frequencies and Percentages of Neurotics and Psychotics Rejecting Rorschach Cards

Number of cards rejected	Neurotics ($N=37$) No.	Neurotics ($N=37$) %	Psychotics ($N=59$) No.	Psychotics ($N=59$) %
0	21	56.8	34	57.6
1	5	13.5	12	20.3
2	4	10.8	7	11.9
3	4	10.8	3	5.1
4	1	2.7	1	1.7
5	2	5.4	2	3.4
6–10	0	0	0	0

Note.—The Kolmogorov-Smirnov two-sample test shows no significant difference at the .05 level between the two percentage distributions.

Table 2

Frequencies and Percentages of Neurotics and Psychotics Rejecting Individual Rorschach Cards

Rorschach Card	Neurotics ($N=37$) No.	Neurotics ($N=37$) %	Psychotics ($N=59$) No.	Psychotics ($N=59$) %	Total No.	Total %
I	0	0	3	5.1	3	3.1
II	6	16.2	3	5.1	9	9.4
III	2	5.4	1	1.7	3	3.1
IV	3	8.1	5	8.5	8	8.3
V	1	2.7	4	6.8	5	5.2
VI	3	8.1	8	13.6	11	11.5
VII	10	27.0	6	10.2	16	16.7
VIII	2	5.4	3	5.1	5	5.2
IX	7	18.9	9	15.2	16	16.7
X	5	13.5	7	11.9	12	12.5

Note.—Critical ratio tests for the significance of the difference between percentages showed a significant difference at the .05 level only for Card VII, the CR being 2.06.

Results

None of the chi square tests for the relationship between Rorschach card rejection and the various personality traits reached an acceptable level of statistical significance, nor were any trends even apparent. Of the neurotic subsample ($N = 37$), 43.2% gave one or more rejections, and of the psychotics ($N = 59$), 42.4% did so. Table 1 shows the frequencies and percentages of subjects from each group rejecting specific numbers of cards. The Kolmogorov-Smirnov two-sample test for differences between the two percentage distributions failed to reach the .05 level of significance.

Table 2 shows the frequencies and percentages of neurotics and psychotics rejecting each individual Rorschach card. Tests of the significance of percentage differences for each card showed that only for Card VII was the difference significant at the .05 level, the critical ratio being 2.03. The significance of this finding vanishes, however, since one out of ten computed statistics attaining the .05 level could result from chance.

Discussion

This attempt to identify personality traits associated with the tendency to reject Rorschach cards failed to find significant relationships. Although the omission of MMPI items was thought to be the "objective" test counterpart of Rorschach rejections, this belief was unsupported by the data. In agreement with other studies (4, 9), it was found that nosological groups (neurotics and psychotics) rejected essentially the same number of cards with comparable frequencies for given numbers of cards and for specific cards. The very close consistency with the findings of the other empirical studies with respect to the differential rejection rates evoked by particular cards is illustrated by Table 3. This table compares the independently obtained orders of the five most frequently rejected cards, arranged from highest to lowest in rate of rejection.

It may be seen that among the five most frequently rejected cards, Nos. IX, VI, X, and VII are common to the findings of all three investigators. It is of considerable interest to note that Sisson's normal group is also included in this communality. Closely related to these stable differential rejection rates are the reported observations of initial reaction times. Phillips and Smith (5), presenting data from 59 normal males ranging in age from 20 to 40 years, showed the following order of the five cards eliciting the longest response latency: VI, IX, II, X, and IV. Similarly, Beck et al. (2) showed the order which they obtained to be: IX, VI, X, IV, and II. The overlap between the five most rejected cards and the five cards eliciting the longest response latency is striking enough to suggest a common determining factor. Although the identification of this factor cannot be established on the basis of the available data, a likely possibility would be that specific characteristics of these particular cards make for difficulty in deriving concepts from them. It is a matter of conjecture, of course, whether these characteristics pertain to the structural properties or the emotional evocativeness of the cards. While it may be seen that the degrees of difficulty in responding to the cards are the same for different nosological groups, comparison with normative data suggests that psychiatric patients experience more difficulty than do normals.

Summary

The Rorschach literature contains considerable inconsistency among statements concerning the incidence and determinants of Rorschach card rejection within the various nosological groups. While empirical studies have elucidated the actual incidence of rejection, there has been no investigation of possible personality variables underlying it. This study sought to explore the question using measures of personality variables derived from the MMPI and the variable of psychiatric diagnosis in a group of 109 hospital patients

Table 3

Comparison of the Orders of the Five Most Frequently Rejected Cards Derived from Empirical Studies

Study	Subjects	Order from highest to lowest				
		1st	2nd	3rd	4th	5th
Tamkin	96 neurotics and psychotics	VII	IX	X	VI	II
Sisson (9)	200 neurotics and schiz.	IX	X	VII	VI	V
Sisson (9)	190 normals	IX	VI	X	V	VII
Mensh (4)	174 neurotics and psychotics	IX	VI	X	IV	VII

containing subsamples of 37 neurotics and 59 psychotics who were predominantly schizophrenic. No relationships were found between Rorschach card rejection and the variables of defensiveness, psychasthenia, depression, pathology, and diagnosis. The five most frequently rejected cards were shown to be in close agreement with the findings of other empirical studies. They were found to overlap to a large extent the cards evoking the longest response latencies, a phenomenon presumably a function of the relative difficulties of the cards in permitting the derivation of concepts. The available data did not permit a decision about whether the difficulty was derived from the structural properties or the emotional evocativeness of the cards. It was shown, however, that the degrees of difficulty in responding to the cards are the same for different nosological groups, and comparison with normative data suggested that psychiatric patients experience more difficulty than do normals.

Received November 11, 1957.

References

1. Beck, S. J. *Rorschach's Test.* Vol. III. In *Advances in interpretation.* New York: Grune & Stratton, 1952.
2. Beck, S. J., Rabin, A. I., Thiesen, W. G., Molish, H., & Thetford, W. N. The normal personality as projected in the Rorschach Test. *J. Psychol.,* 1950, **30**, 241–298.
3. Klopfer, B., & Kelley, D. *The Rorschach technique.* Yonkers, N. Y.: World Book, 1946.
4. Mensh, I. N., & Matarazzo, J. D. Rorschach card rejection in psychodiagnosis. *J. consult. Psychol.,* 1954, **18**, 271–275.
5. Phillips, L., & Smith, J. G. *Rorschach interpretation: Advanced technique.* New York: Grune & Stratton, 1953.
6. Rorschach, H. *Psychodiagnostics.* (3rd Ed.) New York: Grune & Stratton, 1942.
7. Scherer, I. W. Prognoses and psychological scores in electroconvulsive therapy, psychosurgery, and spontaneous remission. *Amer. J. Psychiat.,* 1951, **107**, 926–931.
8. Scherer, I. W., Winne, J. F., Clancy, D. D., & Baker, R. W. Psychological changes during the first year following prefrontal lobotomy. *Psychol. Monogr.,* 1953, **67**, No. 7 (Whole No. 357).
9. Sisson, B. D., Taulbee, E. S., & Gaston, C. O. Rorschach card rejection in normal and psychiatric groups. *J. clin. Psychol.,* 1956, **12**, 85–88.
10. Tamkin, A. S., & Scherer, I. W. What is measured by the "Cannot Say" scale of the group MMPI? *J. consult. Psychol.,* 1957, **21**, 370.

Journal of Consulting Psychology
Vol. 22, No. 6, 1958

A Validity Study of the Autonomy and Deference Subscales of the EPPS

Darrell Gisvold[1]
Montana State University

The validity of a test or of a personality inventory is frequently defined as the extent to which it actually measures what it purports to measure. However, criterion measures of validity in which data are collected in actual behavioral situations are generally not available. With the exception of a recent study by Bernadin and Jessor (2), the correlations obtained thus far are between the Edwards Personal Preference Schedule and self-ratings or ratings by peers, and other personality inventories.

In the study by Bernadin and Jessor, the construct of Dependency was defined in order to predict the behavior in Asch's experimental situation from behavior on the Autonomy and Deference subscales of the EPPS. In the study, the authors employed Asch's technique of measuring group behavior as the criterion for one of the propositions in the construct, Dependency. They suggested that the reason for the failure to find a correlation between Asch's behavior and behavior on the EPPS was due to the fact that the subjects were placed in a concrete reality situation.

In the present study, in order to relate test behavior with the criterion, the term *conformity behavior* has been defined as the need not to be different, to follow the opinions and suggestions of others, to conform to the group. The conformist has less ability to tolerate his own impulses, less originality, and a greater tendency to follow external and socially approved values.

According to Edwards, the Personal Preference Schedule is a measure of an individual's needs or motives, which college students and adults seek to satisfy in the daily conduct of their lives. The needs measured by the EPPS were drawn from the studies of Murray (5). Edwards' description of the Autonomy and Deference subscales of the EPPS is as follows:

> Autonomy: To be able to come and go as desired, to say what one thinks about things, to be independent of others in making decisions, to feel free to do what one wants, to do things that are unconventional, to avoid situations where one is expected to conform, to do things without regard to what others may think, to criticize those in positions of authority, to avoid responsibilities and obligations.
>
> Deference: To get suggestions from others, to find out what others think, to follow instructions and do what is expected, to praise others, to tell others that they have done a good job, to accept the leadership of others, to read about great men, to conform to custom and avoid the unconventional, to let others make decisions (4, p. 5).

The Autonomy subscale purportedly measures the same personality variable, conformity behavior, as defined above. Furthermore, there is indication that the Deference subscale is measuring a trait which is in opposition to the trait measured by the Autonomy subscale.

Procedure

The purpose of this study was to determine the validity of the Autonomy and Deference subscales of the EPPS by correlation with a criterion measure of conformity behavior. The method of measuring conformity behavior is a modification of Asch's basic group situation developed for measuring conformity behavior.

Asch's (1) technique of measuring conformity behavior requires the subjects (*S*s) to report which of three parallel lines is equal in length to a standard line. One naive *S* is placed in a group with three other *S*s who have been previously instructed to give incorrect answers. Asch found that an individual is not likely to deviate from the group; he is more likely to go against the evidence of his own senses when the difference is most difficult to perceive, when the group is unanimous against him, and when those in opposition increase in numbers.

---

[1] Now at Creighton University Medical School.

The present study employed Asch's method, with the exception that four naive *S*s were used, as in Crutchfield's (3) modification. The present study did not employ Crutchfield's electrical apparatus to record responses.

In the present study, 50 volunteer *S*s, 31 males and 19 females, were drawn from four sections of an elementary psychology course. An attempt was made to control prestige effects by selecting not more than two *S*s from a class to participate in each group of four *S*s. Four *S*s were instructed to enter the experimental room and take seats randomly among four partitioned chairs. The only person they could see was the experimenter, who was seated about 13 ft. in front of them at a small table. At each *S*'s desk was an instruction sheet and three response cards numbered 1, 2, and 3. In each of the 18 trials, the experimenter displayed two large cards: the card on the left showed the standard line, and the card on the right showed three parallel comparison lines numbered consecutively 1, 2, and 3. Then the four *S*s made their judgments to the standard by raising a card numbered to correspond with the selected comparison line.

The illusion of the procedure rested on each *S*'s belief that he was the fourth and final person to make a judgment. When the experimenter reported the judgments of the first three *S*s, he was creating a false group majority. Actually, no one had reported. The pressure to conform to the group was artificially produced. False majority judgments were reported in 9 of the 18 trials. Whether the experimenter reported a true or false majority response, it was unanimous; the first three *S*s were reported to have given the same judgment.

The differences to be discriminated were considerable; most unequal comparison lines were clearly longer or shorter than the standard. The comparison lines differed from the standard by varying amounts between .5 and 1.75 inches; no attempt was made to maintain a constant ratio between them. On successive trials the equal line appeared randomly in different positions.

After being subjected to the group conformity situation, the group of 50 students was given the Edwards Personal Preference Schedule within a two-week period.

Results

A comparison of the data obtained from the students sampled in this study was made with both the results obtained by Asch in his measurement of group conformity and that of Edwards' normative sample of the Personal Preference Schedule.

A control study was made in order to determine whether the *S*s' conformity errors were due to indiscriminable line lengths. In the control group of 10 *S*s, an average of .8 errors were made in the 9 critical trials. The expected frequency of error on a chance basis should have been 67%. The frequency of error found was 9%. For the sample of 50 *S*s used in this study, it was found that there were 32% conformity errors. This approximates both the results of Asch and Crutchfield of about 30% conformity errors.

An analysis of the difference between the subscale variances of the present sample and Edwards' Normative sample was made by employing the variance ratio, F. In all cases, the F ratios computed were not large enough to reject the null hypothesis at the .05 level of confidence. Further analysis was made by testing the subscale mean difference of the two samples. A t test was employed to test the mean differences. The only significant differences found between the two population samples was on the Succorance subscale and the consistency scores. The former was significant at the .01 level of confidence, and the latter was significant at the .05 level in a direction toward greater consistency of the present sample.

The product-moment correlation between the scores on the Autonomy subscale and conformity scores from the group conformity situation was found to be −.54, significant at the .02 level of confidence. Therefore, there is a high degree of assurance that the Autonomy subscale is measuring the need for Autonomy as described by Edwards.

The product-moment correlation between the scores on the Deference subscale and conformity scores from the group conformity situation was found to be .17; this value was not large enough to reject the null hypothesis at the .05 level of confidence. The correlation ratio eta was computed in order to test for the possible existence of a curvilinear relationship between regression lines, and no significant difference was found.

Summary

The aim of this study was to determine the empirical validity of the Autonomy and Deference subscales of the EPPS. A group situation developed by Asch to measure conformity behavior was used as the criterion.

The correlation between the conformity scores and the scores on the Autonomy subscale was found to be −.54, significant at the .02 level of confidence. Therefore, the Autonomy subscale of the EPPS as described by Edwards is empirically valid with respect to the criterion of conformity behavior as developed for this study.

The correlation between the conformity scores and the scores on the Deference subscale was not significant. On the basis of the criterion measure, conformity behavior, the Deference subscale does not predict an individual's conformity behavior. These findings indicate that a person who has a need for Deference does not necessarily exhibit an equal need to conform to group situations.

Received December 12, 1957.

References

1. Asch, S. E. *Social psychology*. New York: Prentice-Hall, 1952.
2. Bernadin, A. C., & Jessor, R. A construct validation of the Edwards Personal Preference Schedule with respect to dependency. *J. consult. Psychol.*, 1957, 21, 63–67.
3. Crutchfield, R. S. Conformity and character. *Amer. Psychologist*, 1955, 10, 191–198.
4. Edwards, A. L. *Edwards Personal Preference Schedule, Manual*. New York: Psychological Corp., 1954.
5. Murray, H. A. *Exploration in personality*. New York: Oxford Univer. Press, 1938.

Journal of Consulting Psychology
Vol. 22, No. 6, 1958

Relationship Between Achievement Motivation Scores and Manifest Anxiety Scores

Donald H. Kausler and E. Philip Trapp
University of Arkansas

A number of studies (e.g., 2, 4, 6, 7) have indicated that task performance may be a function of achievement motivation. In these studies, however, no effort was made to control for other motivational or drive variables. Apparently, the assumption was taken that the other drive variables operate randomly in groups differentiated on measures of achievement motivation.

This assumption is questioned by the present authors. Measures of achievement, which are determined by test scores, may, in effect, be the product of a number of motivational factors, achievement motivation being but one. For example, performance on an achievement test possibly could be related to the anxiety drive level of subjects. Groups formed on the basis of achievement test scores would then also represent groups differentiated with respect to anxiety drive level.

The present empirical study is designed to investigate this conceivable relationship between achievement motivation and manifest anxiety drive. More specifically, the study involves a comparison between two sets of scores —one set on an achievement test and the other on a test of manifest anxiety drive. The null hypothesis as involved in this study states that there is no relationship between level of achievement motivation and level of manifest anxiety drive.

Method

Subjects. One hundred and three male students in general and abnormal psychology courses at the University of Arkansas served as *S*s. Only male *S*s were used because of the strong possibility, as suggested by Atkinson (1), of a pronounced sex difference in achievement motivation. All testing was conducted during regular class periods.

Instruments. The complex motivation test developed by French (4, 5) served as the achievement motivation measure. The test was scored for achievement motivation by the scoring system devised by French. Close agreement was found between test scores for two scorers working independently.

The Taylor Manifest Anxiety Scale (8) served as the measure of anxiety drive level. The standard administration and scoring procedure outlined by Taylor were followed.

Procedure. The achievement motivation test, disguised as a "Test of Insight," was administered early in the semester. The *S*s were instructed that the study was part of a research project designed to develop a new psychological test. Approximately one month later, the anxiety scale was administered. This scale, titled "Biographical Inventory," was described to the *S*s as a source of biographical information for comparing students at Arkansas with students elsewhere. Apparently, the *S*s did not suspect a relationship between the two parts of the study.

Results

A correlation coefficient was computed between the scores received on the two tests. Since the scores on both the French test and the Taylor test are ordinal values, a nonparametric statistic, the Spearman rank-difference rho, was used. A rho of −.20 was obtained, which is significant at the .05 level of confidence ($t = 2.05$). Thus, when the entire range of scores is considered, there is a small, but significant, negative correlation between achievement and anxiety test scores.

Table 1

Achievement Motivation and Anxiety Drive Level: Observed and Expected Frequencies

Anxiety drive	Achievement motivation	
	High	Low
High	*5* 2	*5* 8
Low	*7* 10	*7* 4

Note.—$\chi^2 = 6.17$ ($p < .02$).

Since most studies employing achievement motivation and anxiety tests compare *S*s only at the two extremes of the scoring range, an additional analysis of the data was performed. Four groups consisting of the top and bottom quartiles on both tests were determined ($N = 26$ in each quartile). A 2 × 2 contingency table was then established and tested by chi square. In the table, the two groups were high and low achievement motivation, and the two classifications were high and low anxiety drive. The results of this analysis are given in Table 1.

An inspection of Table 1 shows that high achievement motivation scores are related to low anxiety scores and that low achievement motivation scores are related to high anxiety scores. The over-all χ^2 is 6.17, which is significant at the .02 level of confidence. When only extreme groups are considered, the negative relationship between the two variables becomes more pronounced. This lends additional support for the rejection of the null hypothesis of no relationship between the two sets of scores.

Discussion

Research, notably from the Iowa group (3, 9), has demonstrated that performance level on various tasks may be affected by anxiety drive level. Consequently, the relationship between achievement motivation and anxiety drive should definitely be taken into account in studies attempting to relate achievement motivation to performance. Otherwise, the effect of different levels of achievement motivation on performance might be contaminated by corresponding systematic differences in anxiety drive level. In other words, an attempt should be made to partial out differences in anxiety drive level when *S*s are differentiated on the basis of achievement test scores.

Two alternative interpretations are offered for the correlation between achievement and anxiety test scores. The first suggests that both tests are measures of the same general drive state. In this case, the correlation simply reflects the amount of overlap. The second interpretation suggests that the two tests are measures of two different but related drive components. In this case, the correlation simply reflects the degree of interaction. The resolution of these opposing explanations is fundamentally dependent upon further validity studies on both achievement motivation and manifest anxiety tests. Current evidence is in the direction of favoring the second interpretation, but the lack of sufficient validation evidence indicates the need for further investigation into the nature of the drive components measured by tests of human motivation.

Summary

A comparison was made between the scores received on the French Achievement Motivation Test and the Taylor Manifest Anxiety Scale by 103 undergraduate male students. A rho of −.20 was obtained ($p < .05$). In addition, a chi square analysis of the data was performed. In the resultant four-fold contingency table, the two groups consisted of the top and bottom quartiles on the achievement motivation test, and the two classifications consisted of the top and bottom quartiles on the anxiety test. The over-all χ^2 was 6.17 ($p < .02$). The results clearly reflect a significantly negative relationship between the two sets of scores, and indicate the need to partial out the anxiety drive component in studies designed to measure the effects of achievement motivation. The question of whether the Taylor scale and the French test are actually measuring two aspects of the same drive state or aspects of two interrelated drive states is raised.

Received December 23, 1957.

References

1. Atkinson, J. W. Explorations using imaginative thought to assess the strength of human motives. In *Nebraska symposium on motivation,* 1954. Lincoln: Univer. Nebraska Press, 1954.
2. Atkinson, J. W., & Reitman, W. R. Performance as a function of motive strength and expectancy of goal-attainment. *J. abnorm. soc. Psychol.,* 1956, **53**, 361–366.
3. Farber, I. E. The role of motivation in verbal learning and performance. *Psychol. Bull.,* 1955, **52**, 311–327.
4. French, Elizabeth G. Some characteristics of achievement motivation. *J. exp. Psychol.,* 1955, **50**, 232–236.
5. French Elizabeth G. Development of a measure of complex motivation. *USAF Pers. Train. Res. Cent., Res. Rep.,* 1956, AFPTRC–TN-56-48.
6. Karolchuck, Patricia A., & Worell, L. Achievement motivation and learning. *J. abnorm. soc. Psychol.,* 1956, **53**, 255–257.
7. Lowell, E. L. The effect of need for achievement on learning and performance. *J. Psychol.,* 1952, **33**, 31–40.
8. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.,* 1953, **48**, 285–290.
9. Taylor, Janet A. Drive theory and manifest anxiety. *Psychol. Bull.,* 1956, **52**, 311–327.

Journal of Consulting Psychology
Vol. 22, No. 6, 1958

Interrelationships Among Three Measures of Need Achievement[1]

Philip Himelstein[2]
University of Arkansas

Arthur E. Eschenbach and A. Carp[3]
Personnel Laboratory, Wright Air Development Center

In recent years, there has been considerable effort devoted to investigating the characteristics of the achievement motive (n-Ach). For this purpose, three tests have been presented which are suitable for research involving large groups: the n-Ach scale of the Edwards Personal Preference Scale (2), French's Test of Insight (3, 4), and McClelland's n-Ach (8). The latter two tests are projective tests of n-Ach, while the Personal Preference Schedule is an objective, forced-choice test which provides, among others, a scale of achievement. In a recent study, Bendig (1) reported a correlation of .11 between McClelland's projective measure of n-Ach and Edwards' objective n-Ach scale. The obtained correlation, with $N = 244$, was not significant at the .05 level. The present study is an attempt to verify Bendig's findings and, at the same time, investigate the relationship between two projective tests of n-Ach, McClelland's and French's.

Procedure

The *S*s in this study were 298 members of an entering class of the Air Force Academy, Denver, Colorado. The *S*s were tested in July, 1957, within one or two days after reporting to the academy. The three tests were administered within a period of two days as part of a larger testing program.

Pictures A, B, D, and E of the McClelland n-Ach test (8, p. 375) were used in this study. All stories were independently scored by two psychologists after practice with stories in the manual. Before scoring the papers obtained from the present sample, the two scorers compared their scores with those of the authors of the McClelland scoring system (8, pp. 335–374). Correlations in the .80's were obtained for the scorers when compared with the manual. This is similar to the level of agreement obtained by other users of this scoring system (6). All 298 papers were scored by both raters. In addition to the independent scores obtained by the two psychologists, a "joint score" was obtained. This score was obtained as follows: Differences of three points or less in total score were averaged; larger differences were settled by rescoring the stories jointly. Interscorer reliability before the scoring conference was .714 for the entire sample.

Form I of French's Test of Insight (4, p. 5) was used in this study. The scoring system for this test is similar in many respects to McClelland's system. This test was scored after the raters had obtained considerable practice in scoring the McClelland n-Ach test. The raters independently scored two intact testing groups ($N = 77$). Rater reliability for these cases was .70. Each scorer then scored approximately half of the remaining cases. The "joint score" in this instance was obtained for the first 77 cases, using the same procedure as described above for the McClelland joint score.

Results and Discussion

It can be seen from Table 1 that no significant relationship was obtained among the three measures of n-Ach. The results bear out Bendig's finding of a lack of relationship be-

---

[1] This investigation was carried out under the Air Force Personnel and Training Research Center in support of Project No. 7719. Permission is granted for reproduction, translation, publication, and use or disposal in whole or in part by or for the United States Government.

[2] With the Personnel Laboratory, WADC, when this study was conducted.

[3] The authors would like to express appreciation to Raymond E. Christal, Eli S. Flyer, John D. Krumboltz, and Ernest C. Tupes, for their assistance in planning the study and reading portions of the manuscript.

Table 1

Intercorrelations of the Measures of Need Achievement

	Tests		
Raters	French-McClelland	French-Edwards	McClelland-Edwards
A	−.024 (172)	.105 (172)	−.018 (298)
B	.047 (203)	.082 (207)	.010 (298)
Joint	−.067 (77)	.019 (77)	.003 (298)

Note.—Numbers in parentheses refer to size of sample.

tween an objective and projective measure of n-Ach. In addition, the present study failed to find a significant relationship between two projective measures of n-Ach. In view of the fairly high interscorer reliabilities for tests of this nature, one would expect, if they are actually measuring a "general need for achievement," that a significant relationship would exist between them.

There are no published reports of the test-retest reliability of the Test of Insight. However, in view of the extremely low test-retest reliabilities reported for the McClelland n-Ach test for one-week (7) and nine-week (6) intervals, it is possible that lack of stability is characteristic of projective tests of need achievement. Thus, the lack of relationship of these measures with each other may be due to a lack of reliability. Another possibility, of course, is that some or all of these measures lack validity.

It should be pointed out that, although this study indicates that these measures are unrelated, their "validity" was not investigated. It is possible that the three scores are measures of three independent but real traits whose only similarity is the names assigned them. This would indicate that some redefining is necessary. At any rate, it should be pointed out, as others (5) have done, that it is dangerous to use interchangeably traits that are similarly defined but measured by different instruments.

Summary

The problem under investigation concerned relationships existing between three measures of achievement motivation: the n-Ach scale of the Edwards Personal Preference Schedule, McClelland's test, and French's Test of Insight. Fairly high interscorer reliabilities were obtained for the two projective tests of achievement. However, no significant relationships were found for the three measures purporting to measure achievement motivation.

Received December 16, 1957.

References

1. Bendig, A. W. Manifest anxiety and projective and objective measures of need achievement. *J. consult. Psychol.*, 1957, **21**, 354.
2. Edwards, A. L. *Manual for Edwards' Personal Preference Schedule*. New York: Psychological Corp., 1954.
3. French, Elizabeth G. Some characteristics of achievement motivation. *J. exp. Psychol.*, 1955, **50**, 232–236.
4. French, Elizabeth G. *Development of a measure of complex motivation. USAF Pers. Train. Res. Cent., Res. Rep.*, 1956, No. AFPTRC-TN-56-48.
5. Iscoe, I., & Lucier, O. A comparison of the Revised Allport-Vernon Scale of Values (1951) and the Kuder Preference Record (Personal). *J. appl. Psychol.*, 1953, **37**, 195–196.
6. Krumboltz, J. D., & Farquhar, W. W. Reliability and validity of the n-achievement test. *J. consult. Psychol.*, 1957, **21**, 226–228.
7. McClelland, D. C. (Ed.) *Studies in motivation.* New York: Appleton-Century-Crofts, 1955.
8. McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

Journal of Consulting Psychology
Vol. 22, No. 6, 1958

Differentiation of Diagnostic Groups by Individual MMPI Scales[1]

Albert Rosen
University of Maryland

A multiscaled instrument such as the MMPI is generally interpreted in terms of configurations of scale patterns and elevations. Information concerning the concurrent validity of individual scales can nevertheless be of considerable value. Besides indicating the degree of differentiation possible between diagnostic groups by means of individual scales, such information can also contribute to the enrichment of configural kinds of interpretations in the psychiatric setting. Of even broader significance is the fact that data on individual scale discrimination can add to our total comprehension of the meanings inherent in a specific scale and therefore are apt to suggest other significant relationships and inferences.

Although the present study was part of a larger project and not planned as a replication, the findings are of special interest because of another report that virtually no differentiation between psychiatric groups is possible using individual MMPI scales. Rubin (6, 7) has reported that the MMPI cannot significantly differentiate between diagnostic groups in a VA psychiatric hospital. He selected at random 93 MMPI records of patients, the total group consisting of 8 chronic alcoholics without psychosis, 24 psychopaths, 33 schizophrenics, and 28 psychoneurotics. Using the analysis of variance method, he found that only one clinical scale, *Sc*, differentiated in some manner between the four diagnostic groups. This paper presents additional data on the discriminating power of the individual MMPI scales, as well as some possible explanations of Rubin's negative results.

Procedure

Five groups from the psychiatry service of the Minneapolis VA Hospital were chosen which were composed of 307 patients diagnosed according to the Veterans Administration psychiatric nomenclature as follows: (*a*) anxiety reaction (anxiety state)—83 cases; (*b*) conversion reaction (conversion hysteria)—49 cases; (*c*) depressive reaction (neurotic depression)—36 cases; (*d*) somatization reaction (psychophysiological reaction, psychosomatic reaction, organ neurosis)—39 cases; and (*e*) schizophrenic reaction, paranoid type—100 cases.

The few published studies providing objective data for the evaluation of diagnostic reliability have emphasized that there is a high degree of inconsistency in psychiatric diagnosis (2, 3, 5). If diagnostic unreliability does prevail, psychological research based upon a diagnostic criterion is apt to be of limited value, for patients classified within any one diagnostic group will be quite heterogeneous, and there will be considerable overlapping between groups.

Because of the importance to the present study of obtaining homogeneous diagnostic groups, only so-called "pure" cases were selected for each sample. Thus, a case was excluded if there was evidence that the patient had been given (*a*) two different diagnoses upon successive admissions, or (*b*) a diagnosis containing a statement of overlapping trends, as for example, "anxiety reaction with paranoid and schizoid trends." There was, however, one exception to this rule. Inasmuch

---

[1] This research was done in the Psychiatry Service of the Minneapolis VA Hospital. The investigator is indebted to Leonard I. Schneider for suggestions leading to improved organization and clarity of the report.

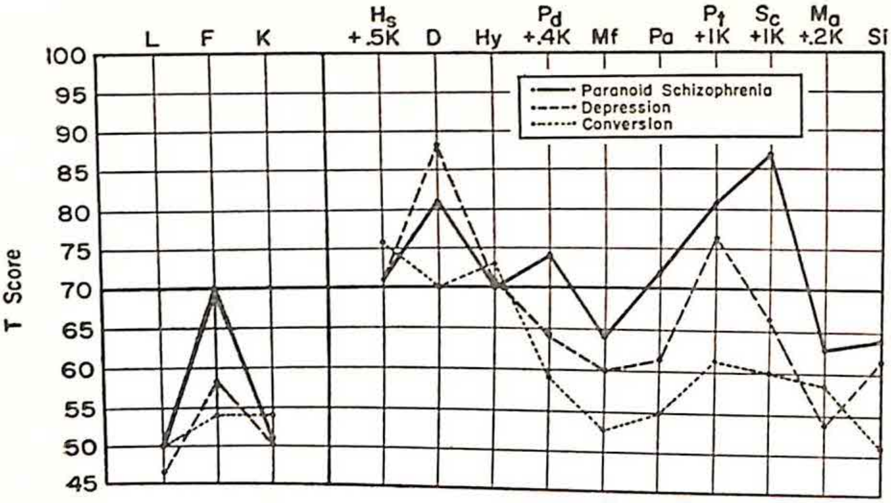


Fig. 1. Mean *T* scores for three diagnostic groups on each MMPI scale.

as a large number of anxiety reaction cases suffer from depression, it was not feasible to exclude cases of "anxiety reaction with depression."

The procedure in selecting patients for the five diagnostic groups was as follows:

1. From a file listing the final diagnoses of all patients who had been tested, the names were selected of "pure" cases who had been diagnosed in one of the five categories listed above.[2]
2. For the purpose of increasing the homogeneity of each diagnostic group, there were included in the study only male, white veterans who had been tested within 15 days after admission and who had not received insulin or electroshock therapy prior to the administration of the MMPI. The median number of days between admission and MMPI administration for *each* of the five final samples was between 3 and 4.
3. The MMPI records of the patients selected were obtained from their test folders, and a few additional cases were removed on the basis of extremely high *L, F, K,* and *?* scores. In the final samples, the median number of *?* items for *each* of the five diagnostic groups was 2–4, Q_3 was 6–14, and only 15 records contained more than 30 *?* items.

Results

The following procedure was used to test the significance of the differences in means of each MMPI scale for the five diagnostic groups:

1. The Bartlett test (4, p. 97) was applied and heterogeneity of variances ($p \leqslant .05$) of the five groups was found for six of the scales: F, K, $Pd + .4K$, Pa, $Sc + 1K$, and Si.
2. For each of these six scales, critical ratios were calculated for the 10 possible differences between means.
3. For each of the other seven scales with homogeneous variances, an analysis of variance test was made.

The mean scores of the diagnostic groups for each MMPI scale are shown in Fig. 1 and listed in Table 1. These *T* scores are rounded numbers because all calculations were made with raw score means and variances. The *F* ratio is also recorded in Table 1 for the seven scales for which this statistic was computed. Since the curves for the Depression and Anxiety groups and for the Somatization and Conversion groups are almost identical, the median profiles for only three groups are shown in Fig. 1 in order to avoid a confusion of lines. On seven of the scales there is an orderly hierarchy of mean scores, with the Paranoid Schizophrenia group being highest, and the Depression, Anxiety, Somatization, and Conversion groups following in order. These seven scales are F, $Pd + .4K$, Mf, Pa, $Pt + 1K$, $Sc + 1K$, and Si. On the D scale, this order for the five groups is the same except for a reversal of the Depression and Paranoid Schizophrenia groups. In the profiles of the Somatization and Conversion groups, the $Hs + .5K$, D, and Hy scores are most elevated with D slightly lower than the other two. This is the so-called "V profile," frequently noted for individual cases diagnosed as conversion or somatization reaction.

---

[2] An exception was the group of patients diagnosed paranoid schizophrenia. Because of their abundance, 100 cases were selected from those most recently discharged.

Among the 130 possible differences in mean scale scores for the five diagnostic groups, 61 are significant at or beyond the .05 level, as listed in Table 2. Eleven such differences would be expected by chance if the statisti-

Table 1

Mean T Scores of Five Diagnostic Groups on Each MMPI Scale

MMPI Scale	Diagnostic groups: Paranoid schizophrenia ($N = 100$)	Depression ($N = 36$)	Anxiety ($N = 83$)	Somatization ($N = 39$)	Conversion ($N = 49$)	F Ratio
L	50	47	48	52	50	2.78*
F	70	58	58	54	54	
K	51	50	51	56	54	
Hs + .5*K*	71	71	73	79	76	2.33
D	81	88	81	71	70	8.94**
Hy	70	71	72	72	73	<1.00
Pd + .4*K*	74	64	63	59	59	
Mf	64	60	57	55	53	11.87**
Pa	72	62	60	57	55	
Pt + 1*K*	81	77	76	66	62	16.16**
Sc + 1*K*	87	67	67	62	60	
Ma + .2*K*	63	54	55	56	58	6.65**
Si	64	62	61	53	51	

* Significant at .05 level.
** Significant at .001 level.

Table 2

Diagnostic Groups Significantly Different in Mean Scores on Each MMPI Scale

MMPI Scale	Diagnostic groups: Paranoid schizophrenia ($N = 100$)	Depression ($N = 36$)	Anxiety ($N = 83$)	Somatization ($N = 39$)	Conversion ($N = 49$)
Mean score of above groups higher than score of—					
L	A D			A D	
F	D A S C				
K				P A D	
Hs + .5*K*					
D	S C	P A S C	S C		
Hy					
Pd + .4*K*	D A S C	C	C		
Mf	A S C	S C	C		
Pa	D A S C	S C	C		
Pt + 1*K*	A S C	S C	S C		
Sc + 1*K*	D A S C		C		
Ma + .2*K*	A D C S				D
Si	S C	S C	S C		

Note.—Each capital letter in the body of the table denotes a diagnostic group as follows: P—Paranoid Schizophrenia; D—Depression; A—Anxiety; S—Somatization; C—Conversion. The table may be read as follows: There are 61 significant differences ($p \leq .05$) between the five diagnostic groups on the various scales. For example, on the *L* scale, there are four significant differences; the Paranoid Schizophrenia group scores higher than the Anxiety and Depression groups and the Somatization group has a higher mean score than the Anxiety and Depression groups.

cal tests were independent (8). Of these 61 differences, 17 are significant at the .05 level, 8 at the .01 level, 5 at the .001 level, and 31 at or beyond the .0001 level. The Anxiety and Depression groups are significantly different from each other only on the *D* scale, and the Somatization and Conversion groups are similar on every scale. If the Anxiety and Depression groups are combined, and the Somatization and Conversion groups are also treated as one group, each of these two groups is effectively differentiated from the other and from the Paranoid Schizophrenia group. Cross-validation of this finding is necessary, however, for the combining of groups was done after inspection of the data. Neither the *Hs* + .5*K* nor the *Hy* scale differentiates the Somatization or Conversion groups from the other three groups.

Discussion

The positive findings of the present study contrast sharply with the almost complete lack of MMPI differentiation between Rubin's diagnostic groups. Three possible explanations of the divergent results will be discussed as follows: (*a*) High *?* scores in Rubin's records; (*b*) criterion contamination in the present study; and (*c*) criterion unreliability in Rubin's groups.

High ? scores. Rubin (7) reported that 9 of the 93 MMPI records in his study contained more than 100 unanswered items. The writer has seen the complete distribution of *? T* scores [3] for the 93 records, and found that 34, or 37% of the records, contain more than 30 unanswered items. Of these 34, 16 have 31–50 such items, 7 have 51–90, 5 have 91–110, and 6 have more than 130 *?* items.

Rubin also indicated that 21% of his records of psychotics had more than 100 unanswered items, whereas only 7% of the neurotics' records and none of the records of psychopaths had more than 100 *?* items. Moreover, the mean *?* scores of Rubin's psychotic, neurotic, and psychopathic groups are also in this descending order. The effect of a large number of unanswered items in any record is to sharply reduce the scores on the MMPI scales. If the number of unanswered items is highest for a group of psychotic patients, next highest for neurotics and still lower for psychopaths, the scores of the MMPI scales of the psychotic and neurotic groups might be reduced to the extent that these two groups could not be distinguished from each other or from the psychopathic group. This phenomenon has apparently occurred in Rubin's study, and could account in large part for his negative results.[4, 5]

Contamination of the criterion. The psychiatrists who made the final diagnoses of the cases included in this study had access to the psychologists' interpretations of the MMPI records. It should be noted, however, that the MMPI interpretations were made in terms of the scale configurations, and therefore should have had minimal effect on the differentiating power of any individual scale. A further mitigating feature is that the psychiatrists claimed they were influenced by the MMPI interpretations primarily in the diagnosis of "borderline" cases. Nevertheless, the criterion contamination, although unavoidable, presents a serious limitation on the evaluation of the results because its effects cannot be precisely measured. Replication of this investigation in a clinical setting where MMPI results do not contribute to final diagnosis would be desirable.

Diagnostic unreliability. In studies in which a diagnostic criterion is used, divergent results may often partly be attributed to the nonhomogeneity of diagnostic categories and to the resultant inconsistency of diagnoses. Any investigation based on highly unreliable criteria is doomed to negative or inconclusive outcomes. Rubin did not mention any attempt at diagnostic purification, and, in fact, the category of alcoholism, which Aaronson and Welsh (1) have indicated, generally overlaps considerably with the psychopathic and neurotic categories.

---

[3] The distribution of *?* scores was obtained through the courtesy of Paul E. Meehl.

[4] Aaronson and Welsh (1) have also discussed the possible effects of extremely high *?* scores in Rubin's study.

[5] As a result of the misleading and indeterminable effect of high *?* scores, it has been the practice at the Minneapolis VA Hospital for the past six years to discourage patients from making *any ?* responses. The Cannot Say cards have been removed from MMPI boxes to implement this practice, so that few records contain any unanswered items. The statistics reported here were computed in 1952.

The null hypotheses which were tested in Rubin's study and in the present investigation can now be restated in order to evaluate further the divergent results and to emphasize the kinds of decisions which might be made in future research into the differentiability of diagnostic groups by means of the individual MMPI scales.

Rubin's null hypothesis may be formulated as follows: Individual MMPI scales do not differentiate the broad categories of neurosis, psychosis, psychopathic personality, and alcoholism without psychosis if the patients are selected without regard to diagnostic purity and to the number of unanswered items in their MMPI records. It is a contribution to our knowledge of the effectiveness of the MMPI to learn that Rubin's null hypothesis cannot be rejected when the MMPI is used in the manner which he employed. It remained to be determined, however, how effective the MMPI is in differentiating diagnostic groups when the size of the *?* scores is controlled and a reasonable degree of homogeneity is present in the criterion groups.

The null hypothesis tested in the present investigation may be stated as follows: Individual MMPI scales do not differentiate between groups of patients diagnosed in the categories of paranoid schizophrenia, neurotic depressive reaction, anxiety reaction, conversion reaction, and somatization reaction if (*a*) "pure" diagnostic groups of patients are selected (with the exception that cases of anxiety reaction with depression are included), (*b*) the number of *?* items in the MMPI records is kept to a minimum, and (*c*) an unmeasurable amount of criterion contamination occurs. It was found that the null hypothesis could substantially be rejected except for the comparison of the Depression and Anxiety groups, exclusive of the *D* scale, and likewise in the comparison of the Conversion and Somatization groups. If, however, the Depression and Anxiety groups are treated as one class, and the Conversion and Somatization groups are combined, each of these two groups is effectively differentiated from the other and from the Paranoid Schizophrenia group.

Summary

A major purpose of this paper was to report an investigation into the effectiveness of the individual MMPI scales in diagnostic group differentiation. A correlated aim was to test the generality of the findings of Rubin who reported that only one MMPI scale differentiated in any manner between four diagnostic groups: psychotics, neurotics, psychopaths, and alcoholics without psychosis.

A total of 307 patients was selected who had been diagnosed in five categories: paranoid schizophrenia, neurotic depressive reaction, anxiety reaction, somatization reaction, and conversion hysteria. An attempt was made to select for each diagnostic category only "pure" cases without overlapping diagnostic trends whose MMPI records contained a minimum of unanswered items. Of 130 possible differences in means for 13 cases, 61 were significant at or beyond the .05 level. The positive findings of the present study, therefore, contrasted sharply with the results obtained by Rubin. Three possible explanations of the divergent results discussed were: large numbers of unanswered MMPI items and diagnostic overlapping in Rubin's groups and criterion contamination in the present study.

Received December 16, 1957.

References

1. Aaronson, B. S., & Welsh, G. S. The MMPI as diagnostic differentiator: A reply to Rubin. *J. consult. Psychol.*, 1950, **14**, 324–326.
2. Ash, P. The reliability of psychiatric diagnoses. *J. abnorm. soc. Psychol.*, 1949, **44**, 272–276.
3. Boisen, A. T. Types of dementia praecox—A study in psychiatric classification. *Psychiatry*, 1938, **1**, 233–236.
4. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
5. Mehlman, B. The reliability of psychiatric diagnoses. *J. abnorm. soc. Psychol.*, 1952, **47**, 577–578.
6. Rubin, H. The MMPI as a diagnostic aid in a veterans hospital. *J. consult. Psychol.*, 1948, **12**, 251–254.
7. Rubin, H. A note on "reply to Rubin." *J. consult. Psychol.*, 1950, **14**, 327–328.
8. Sakoda, J. M., Cohen, B. H., & Beall, G. Test of significance for a series of statistical tests. *Psychol. Bull.*, 1954, **51**, 172–175.

Journal of Consulting Psychology
Vol. 22, No. 6, 1958

Changes in Responses to a Set of Personality-Inventory Items as a Function of Age[1]

Richard H. Walters
University of Toronto

and Misha S. Zaks
Northwestern University Medical School

Strong's first interest-maturity scale was developed by contrasting the responses of 55-year-old men and 15-year-old boys to items of his Vocational Interest Blank. Mean scores on this scale showed a rapid increase from age 15 to about age 21, then a more gradual rise to about age 50. It seems reasonable to suppose that responses to certain personality-inventory items might change as a function of age and that a scale composed of such items might be sensitive to some aspects of the maturational process. The present research constitutes an inquiry into the feasibility of such a scale.

A 160-item inventory was administered to 32 15-year-old boys and 37 15-year-old girls. The boys' responses to each item were compared with those of 50 adult males, the girls with those of 50 adult females. Twenty-one items with which adolescents agreed significantly more often than the same-sex adults were thus identified.

These items were now treated as a scale, and inventories of *S*s in various age groups were scored for the number of agreements with the items. An analysis of variance showed a highly significant increase in mean scores for high-school boys over the age groups of 16, 17, and 18–19 years ($p < .001$). Mean scores of girls also significantly increased over the available age groups of 15, 16, and 17 years ($p < .02$). A graphical representation of results showed a relatively rapid increase in mean scores during adolescence, and a slower increase during early adult years. Data from male industrial employees suggest that scores may continue to increase slowly up to 50 or 60 years.

Since the total number of agreements to all 160 items did not significantly increase for male *S*s from 15 to 18–19 years, changes in scale scores cannot simply reflect a greater readiness of adolescents to agree with items regardless of their content. The mean scores of 16 high-IQ and 16 low-IQ *S*s were not significantly different. Since these *S*s were matched approximately for age, the scale apparently reflects something more than an increase in mental age.

The meaning of the increase in scores with age is nevertheless quite unclear at the present time. The changes may have no significance whatsoever beyond the fact that they occur. Even so, they constitute a warning to psychologists who administer personality inventories to adolescent *S*s and then interpret them with the aid of adult norms. The question should not, however, be left thus undecided. High-scoring and low-scoring *S*s in the same age range might be compared in their responses to other types of stimuli. If significant correlates of scale scores were thus identified, it might be possible to construct a scale which reflected important aspects of the maturational process.

Brief Report.
Received June 6, 1958.

---

[1] An extended report of this study may be obtained without charge from Richard H. Walters, Department of Psychology, University of Toronto, Toronto 5, Ontario, or for a fee from the American Documentation Institute. Order Document No. 5697, remitting $1.25 for microfilm or $1.25 for photocopies.

Journal of Consulting Psychology
Vol. 22, No. 6, 1958

"Ideal Self" Instructions, MMPI Profile Changes, and the Prediction of Clinical Improvement[1]

Gerald M. Rapaport[2]
Fort Dix, New Jersey

A number of recent studies on the Minnesota Multiphasic Personality Inventory have dealt with the problem of response sets inherent in that test (1, 6, 7) and with the effects of experimentally induced instructional sets on test performance (4, 8, 10, 11). These studies were undertaken to specify those variables which may influence patterns of response apart from the content of the test items and to determine whether these systematic effects can be utilized to improve the diagnostic and prognostic merit of the instrument.

The present study investigated the effects of an "ideal self" instructional set on test performance. The following hypotheses were put forth: (*a*) the introduction of an "ideal self" set produces changes in the MMPI patternings of psychiatric patients relative to the values obtained under the customary instructional procedures; (*b*) the changes in MMPI patterns are in the direction of fewer elevated scale scores; (*c*) the more deviant the original MMPI record, the more marked the change; (*d*) the interindividual variability is significantly less on the "ideal self" pattern than on the "real self" pattern.

A study by Grayson & Olinger (10) closely paralleled the present design. Employing the novel instructional set "typical well-adjusted adult on the outside," the authors found that the ability to improve MMPI performance on the second testing was a favorable prognostic index of early hospital discharge. In the present study, an attempt was made to determine whether Grayson & Olinger's prognostic finding held true under the present instructional procedures.

Tests of the Hypotheses

Procedure

The sample used in the present study consisted of 48 military personnel, all of whom were psychiatric patients. The subjects (*S*s) were drawn over a six-month period from the psychiatric facilities of an Army basic training center.[3] Thirty-three *S*s were seen as outpatient referrals, while the remaining 15 were psychiatric inpatients. Diagnostically, the sample consisted of 12 cases of schizophrenia, 11 individuals diagnosed NND (no neuropsychiatric disease), 20 character and behavior problems, two psychotic depressions, and three cases of epilepsy. The MMPI was administered to each *S* twice, first under the customary "real self" instructional set and, immediately thereafter, under the "ideal self" instructions. The patients invariably accepted the second MMPI as part of the routine clinical psychological workup.

Results and Discussion

Table 1 shows the mean for "real self" and "ideal self" administrations of the test. The group's "real self" pattern reflected a good deal of psychiatric distress, while the "ideal self" pattern manifested an absence of clinical pathology. The hypothesis which predicted that the new instructional set would

---

[1] The author wishes to express his gratitude to the following individuals for their aid in data collection, statistical analyses, and for their suggestions: Robert J. Marshall, Arthur Burdett, Bernard Doran, Charles Henin, Alfred Herbert, David Perry, and John Ryan.

[2] Currently at Illinois Institute of Technology.

[3] This reflects an incidence rate of .004 per thousand post population per annum.

Table 1

Means, Standard Deviation, and Tests of Significance of "Real Self" and "Ideal Self" Differences

Scale	Mean Real	Mean Ideal	Mean Change	SE Diff.	*t*	*SD* Real	*SD* Ideal	Diff.	*F*
L	4.87	9.54	4.67	.53	8.86**	2.66	2.84	.18	1.07
F	9.58	5.73	3.85	.70	5.50**	5.55	4.67	.88	1.10
K	13.08	19.56	6.48	.78	8.30**	4.23	4.70	.47	1.11
Hs	20.56	14.31	6.25	1.27	4.13**	8.03	4.36	3.67	1.84*
D	30.41	18.83	11.58	1.54	7.52**	9.52	4.39	5.13	2.17**
Hy	28.21	20.68	7.42	1.13	6.47**	7.64	5.42	2.22	1.41
Pd	27.37	23.25	4.12	.85	4.48**	5.27	4.51	.76	1.17
Mf	26.43	24.54	1.89	.76	2.48*	6.50	4.62	1.88	1.41
Pa	15.00	9.87	5.12	.75	6.83**	5.36	3.56	1.80	1.51
Pt	37.35	25.71	11.64	1.59	7.32**	10.32	5.87	4.45	1.76*
Sc	38.87	28.02	10.85	1.89	5.79**	12.96	6.98	5.98	1.86*
Ma	23.02	23.04	.02	.83	.02	5.19	4.38	.81	1.18

* Significant beyond .05 level.
** Significant beyond .01 level.
Note.—These figures are expressed in MMPI raw scores.

produce statistically significant changes in the over-all MMPI pattern was confirmed as evidenced by 10 *t*'s significant at the .01 level of confidence and one *t* significant at the .05 level. The second hypothesis, which proposed that the change would be in the direction of greater freedom from mental distress was likewise confirmed, as evidenced by the statistically significant downward changes of the *F*, *Hs*, *D*, *Hy*, *Pd*, *Pa*, *Pt*, and *Sc* scales.

The third hypothesis was tested by correlating each individual's "real self" score on each scale with the discrepancy scores (real self minus ideal self.) The Pearson *r*'s are shown in Table 2. Significant relationships, established for each of the nine clinical scales, indicate that the more pathological the "real self" score, the greater the discrepancy between "real self" and "ideal self."

Table 2

Correlations Between Real Self and Extent of Change

Scale	Correlation
Hs	.85*
D	.93*
Hy	.85*
Pd	.68*
Mf	.71*
Pa	.77*
Pt	.84*
Sc	.85*
Ma	.66*

* Significant beyond .01 level of confidence.

The final hypothesis was tested by comparing for each scale the standard deviation of the "real self" distribution with the standard deviation of the "ideal self" patterns. Ten of the twelve MMPI scales shifted in the predicted direction, four of the variability changes (*Hs*, *D*, *Pt*, *Sc*) attaining statistical significance as measured by the *F* test (5, pp. 163–165). The probability of obtaining the present distribution of variability changes by chance is less than .05 as tested by chi square. Therefore, it may be concluded that patients taking the MMPI show a higher degree of correspondence in specifying what they regard as "ideal" behavior than in assessing their present state of psychological adjustment.

In view of the close correspondence between the methodologies of the Grayson-Olinger study (10) and the present investigation, a comparison of the two sets of findings was undertaken. The mean "real self" and novel instructional set profiles for the two samples are presented in Fig. 1. An inspection of the plotted means shows that on the "real self" administration, the present sample manifested fewer psychopathic trends but more neurotic and psychotic features than the Grayson-Olinger sample, while on the novel instructional sets the two profiles were highly similar in shape. It appears that the new instructional

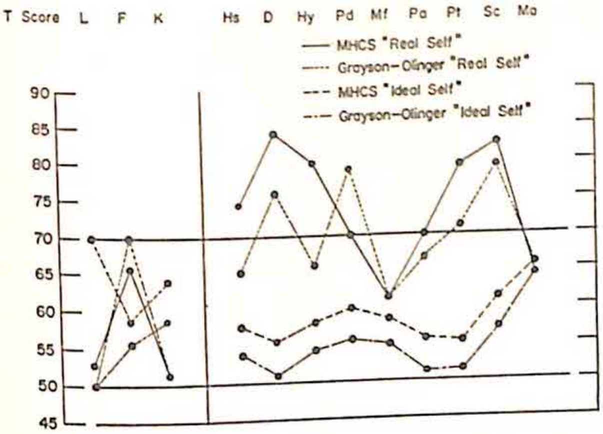


Fig. 1. Comparison of mean profiles of Grayson-Olinger sample with means of MHCS[4] sample.

sets employed in the two studies have exerted virtually the identical effect upon the over-all distributions of the MMPI scales.

Test of the Relationship Between MMPI Improvement and Clinical Prognosis

Procedure

Using an inpatient psychiatric sample, Grayson and Olinger found that MMPI improvability was related to early discharge from the hospital. To determine whether Grayson and Olinger's finding could be replicated in the present data, follow-up data were obtained for *S*s receiving diagnoses other than NND. The obtainable information, available for 32 of 37 *S*s receiving psychiatric diagnoses, consisted of the patient's military status at the time of follow-up. For the purpose of the present evaluation, the *S*s were divided into two groups: those receiving premature discharges either for psychiatric disease or characterological inaptitude or unsuitability; and those successfully pursuing their military tours of duty. Based upon the Grayson-Olinger findings, the following predictions were offered: (*a*) On each MMPI scale the discrepancy scores between "real self" and "ideal self" is greater for the "in service" (good prognosis) group than for the "premature discharge" (poor prognosis) group. (*b*) The total discrepancy score differentiates between the "in service" and "premature" groups.

Results and Discussion

The first hypothesis was tested by *t* tests comparing the means of the discrepancy scores for each scale. While none of the *t*'s attained statistical significance, an interesting trend in line with the prediction was manifested. The means of the discrepancy scores for the "in service" group were of greater value than the corresponding means for the "premature discharge" group on eight of the nine clinical scales. The probability of obtaining the present distribution of discrepancy scores on the basis of chance is .048, which warrants a rejection of the null hypothesis at the .05 level of confidence.

The second hypothesis was tested by comparing the total reduction in *T* scores for *S*s in the good prognosis group, with the reduction in *T* scores for *S*s in the poor prognosis group. The results of this test, which directly paralleled the one undertaken by Grayson and Olinger, are shown in Table 3. In the Grayson-Olinger study, the "good prognosis" group showed a significantly greater percentage of individuals meeting each of the three criterion scores than did the "poor prognosis" group. For the present sample, differences in the predicted direction occurred on two of the three criterion scores. However, the differences did not attain statistical significance as measured by chi square, which indicates that in the military sample, the ability to improve MMPI performance is not related to environmental prognosis.

Table 3

Comparison of the Percentage of Individuals Meeting Three Criterion Scores of MMPI Improvement

	Reduction in MMPI total *T* score		
Sample	90 or more	65 or more	45 or more
Grayson-Olinger			
Good prognosis group	47%	68%	74%
Poor prognosis group	19%	35%	42%
Military			
Good prognosis group	57%	81%	87%
Poor prognosis group	57%	57%	69%

---

[4] Mental Hygiene Consultation Service, Fort Dix.

In comparing the two samples, it can be seen that the present sample as a whole manifested a higher percentage of MMPI improvement at each successive criterion score. Of special significance is the fact that the "poor prognosis" group in the present sample showed approximately as high a percentage of individuals meeting the criterion scores as the "good prognosis" group of the Grayson-Olinger sample. Whether or not the intergroup differences in percentages of individuals meeting improvability criteria are of statistical significance cannot be ascertained, as the breakdown of the number of cases in Grayson-Olinger's group was not indicated. However, another measure of intersample differences in improvability was obtainable. In the Grayson-Olinger sample, 5 of 45 cases (11%) manifested completely normal patterns on their second MMPI in that all nine clinical scores were below the critical range. On the other hand, 15 of the 37 patients receiving psychiatric diagnoses in the present sample (42%) produced completely improved "ideal self" records. A chi square test which compared the two samples with respect to the percentage of complete improvement gave a value of 10.30, which warrants a rejection of the null hypothesis at the .02 level of confidence. This statistically significant finding, along with the between group differences shown in Table 3, indicates that the military sample manifested significantly greater MMPI improvement than the Grayson-Olinger sample. Two explanations of this difference suggested themselves: (*a*) the novel instructional sets employed in the two studies produced essentially different response sets among the *S*s; (*b*) populational differences accounted for the diverse findings.

To determine whether the instructional sets were different, the MMPI was administered according to the Grayson-Olinger methodology to 12 psychiatrically diagnosed military patients. Of these *S*s, nine improved their patterns, and seven manifested complete improvement. Five of the eight inpatients among this group manifested complete improvement. The high percentage of complete improvement corresponds to the data obtained under the "ideal self" instructional set. It therefore appears that populational variables, rather than the differential effects of the instructions, determined the intersample difference in MMPI improvability. The following are among the variables characterizing the military sample. For the most part, the sample consisted of young men in the 20—30 age range, many of whose problems were acute and situational in nature. Moreover, as a number of writers have stated (2, 3, 9), the motivational component in a number of military psychiatric cases differs from that found among civilian patients as a result of more immediate "secondary gain" from the illness. Secondary gain is particularly great during times of stress such as combat or—to a somewhat lesser degree—basic training. It is quite likely that one or more of these variables could account for the fact that MMPI improvability, though high for this sample, does not relate to environmental prognosis. One possible explanation would be the fact that since the motivational level of psychiatric *S*s in this population is at times directed less towards the attainment of the "ideal self" state of asymptomatic behavior than towards maintenance of "real self" symptomatology, a number of these individuals may not actually be striving—at that time—towards attainment of the ideal. In short, the ability to regard a symptom-free psychological state as synonymous with one's ideal may only be a prognostically favorable attribute if there is not the complicating factor of immediate secondary gain from illness.

Whatever the explanation of the failure of the present sample to reproduce the Grayson-Olinger findings, it is suggested that the Grayson-Olinger prognostic finding may be of somewhat restricted generality. Further work on this question is needed in order to clearly delineate the predictive utility of this prognostic tool.

Summary and Conclusions

The present study investigated the influence of an "ideal self" instructional set on MMPI performance. Four hypotheses were tested: (*a*) The introduction of an "ideal self" set produces changes in the MMPI patternings of patients referred for psychiatric consultation; (*b*) these changes are in the direction of fewer scores in the "critical" or "unhealthy" range; (*c*) the more deviant the original MMPI record, the more marked the changes; (*d*) the interindividual variability is significantly less on the "ideal self" pattern than on the "real self" pattern.

The MMPI was administered twice to a group of 48 military psychiatric patients. The first administration of the test was under the usual instructional set while the second was under an "ideal self" set. The above four hypotheses were confirmed.

In a study highly similar to the present investigation, Grayson-Olinger, using a "typical well-adjusted person" instructional set, found that MMPI improvability was related to clinical prognosis. An attempt was made to test the generality of Grayson and Olinger's prognostic finding. Follow-up data were obtained for the majority of military subjects six months to one year following their participation in the study. It was determined that the ability to improve MMPI performance was not significantly related to prognosis, although the trend was in that direction. Moreover, it was noted that the military sample showed a significantly greater over-all ability to improve MMPI performance than did the Grayson-Olinger sample. The possibility that the intersample difference in MMPI improvement could be attributed to instructional differences was ruled out. It therefore appears that populational variables account for the between-group difference. While further study of the Grayson-Olinger prognostic finding is called for, it is suggested that these authors' finding may be of restricted generality.

Received November 26, 1957.

References

1. Barnes, E. H. Factors, response bias, and the MMPI. *J. consult. Psychol.*, 1956, **20**, 419–421.
2. Briskin, G., & Stennis, J. Improving predictability of the MMPI. *U. S. Armed Forces med. J.*, 1957, **8**, 539–543.
3. Bushard, B. L. The U. S. Army's Mental Hygiene Consultation Service. Paper read at Symposium on Preventive and Social Psychiatry, Walter Reed Army Institute of Research, April, 1957.
4. Cofer, C. N., Chance, J., & Judson, A. J. A study of malingering on the MMPI. *J. Psychol.*, 1949, **27**, 491–499.
5. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
6. Fordyce, W. E. Social desirability in the MMPI. *J. consult. Psychol.*, 1956, **20**, 171–175.
7. Fricke, B. G. Response set as a suppressor variable in the OAIS and MMPI. *J. consult. Psychol.*, 1956, **20**, 161–169.
8. Gough, H. G. Simulated patterns on the MMPI. *J. abnorm. soc. Psychol.*, 1947, **42**, 215–225.
9. Gough, H. G. Diagnostic patterns on the Minnesota Multiphasic Personality Inventory. *J. clin. Psychol.*, 1946, **2**, 23–37.
10. Grayson, H. M., & Olinger, L. B. Simulation of "normalcy" by psychiatric patients on the MMPI. *J. consult. Psychol.*, 1957, **21**, 73–77.
11. Rosen, E. Self-appraisal, personal desirability, and perceived social desirability of personality traits. *J. abnorm. soc. Psychol.*, 1956, **52**, 151–158.

CRPM. The scores of the CRPM were slightly extrapolated to give a mental age range from four years six months to eleven years six months.

Means and variances were calculated for the SB scores and the *t* test used to determine the significance of the difference of means. Results were compared on the basis of age, sex, color, color-sex (Negro boys vs. Negro girls, etc.), age-color-sex (seven-year-old Negro boys vs. seven-year-old white boys, etc.), and SB IQ levels (SB below 70, 70–89, etc.). CRPM scores were analyzed in the same manner.

Each child's difference score (SB IQ minus CRPM IQ) was treated as a raw score and the significance of the difference of the means of the differences determined for the above groupings.

Results [1]

Stanford-Binet mean differences were slight and can be attributed to chance for the following groups: Negro vs. white children, boys vs. girls, Negro boys vs. Negro girls, white boys vs. white girls, and age-color-sex groups. There was a tendency for the SB means to decrease as age increased. The difference of SB means for the seven- and eight-year-old children was 2.44 IQ points, with a *t* value of 2.00, which is significant at the .05 but not the .01 level of confidence. The mean of the eight-year-old group was 2.73 points higher than that of the nine-year-old group. The *t* for this difference was 2.17, significant at the .05 level of confidence. However, the decrease in IQ with age was comparable for Negro and white children and for boys and girls. Early school retardation may account, in part, for these differences. The inference is that in this population of children from low socioeconomic level, intelligence as measured by the SB is comparable for color, sex, color-sex, and age-color-sex groups. Hence, the basic assumption is made that there are no real differences in intelligence as measured by the SB for these Negro and white boys and girls.

Significant differences of CRPM IQ means were found when groups were compared on the basis of color. Table 1 shows that the white children had a mean CRPM IQ which was 10.29 points higher than the mean of the Negro children. This difference was significant well beyond the .01 level of confidence, having a *t* value of 8.23. This is in gross contrast to the SB IQ results where no significant differences were found.

When the data were analyzed to determine whether this relationship held for varying SB IQ levels, the results were similar. There were significant differences between the means of Negro and white children for all of the SB IQ groupings (SB below 70, 70–89, 90–109, 110–129, 130 and above). The white children's means were higher than the Negro children's means and the differences were significant at the .01 level of confidence, except for the SB group of 110–129. This latter difference was significant at the .02 level of confidence.

When CRPM IQ mean differences were considered for age-color-sex groups (seven-year-old Negro boys vs. seven-year-old white boys, seven-year-old Negro girls vs. seven-year-old white girls, etc.) the *t* values of the differences were all above 3.00 with the exception of the nine-year-old Negro and white boys. The latter difference had a *t* value of 1.80.

There was no significant difference between the CRPM means of boys and girls nor between Negro boys and girls or white boys and girls. As with the SB IQs, there was a tendency for the CRPM means to decline with age. The mean difference between seven- and eight-year-olds was significant at the .05 level of confidence. The mean of the nine-year-old group was lower than the eight-year-old mean, but the difference was not significant.

The difference scores (each child's SB IQ minus his CRPM IQ) give a picture of the discrepancy between individual IQs on the two tests. It is to be remembered that for the varying groups, the SB IQ means were essentially the same. Here, again, color differences were significant well beyond the .01 level of confidence.

---

[1] Tables of SB and CRPM raw scores and tables of SB, CRPM, and difference score means, variances, and group totals by SB IQ levels have been deposited with the American Documentation Institute. Order Document No. 5739, remitting $1.25 for microfilm or $1.25 for photocopies.

Table 1

Means, Standard Deviations, and t's for SB IQs, CRPM IQs, and Difference Scores by Color, Age, and Sex Groupings

		SB IQ			CRPM IQ			Diff. scores[a]		
Groups	N	Mean	SD	t	Mean	SD	t	Mean	SD	t
Negro	349	90.3	13.2	0.27	80.5	16.3	8.23**	9.8	14.8	9.26**
White	440	90.6	15.1		90.8	18.9		−0.2	15.6	
Boys	389	89.8	13.7	1.25	86.9	18.4	1.05	2.9	16.2	2.32*
Girls	400	91.1	14.7		85.5	18.7		5.6	15.8	
7 Years	271	92.9	13.6	2.00*	88.8	18.3	2.11*	4.1	16.0	0.71
8 Years	273	90.5	14.8		85.3	19.6		5.1	16.5	
8 Years	273	90.5	14.8	2.17*	85.3	19.6	0.61	5.1	16.5	1.24
9 Years	245	87.7	13.9		84.3	17.3		3.4	15.8	
Negro boys	164	90.2	12.8	0.07	81.8	16.7	1.46	8.4	15.6	1.67
Negro girls	185	90.3	13.6		79.3	15.9		11.1	13.9	
White boys	225	89.5	14.4	1.52	90.6	18.7	0.15	−1.1	15.4	1.29
White girls	215	91.7	15.7		90.9	19.2		0.8	15.8	
Negro boys, 7	50	91.9	12.3	0.26	82.8	14.7	3.62**	9.1	14.0	3.99**
White boys, 7	67	92.5	12.2		94.3	19.6		−1.8	15.4	
Negro boys, 8	61	90.9	13.3	0.79	80.5	17.8	3.18**	10.4	16.9	4.08**
White boys, 8	78	89.0	15.0		90.2	17.9		−1.2	16.2	
Negro boys, 9	53	88.0	12.2	0.19	82.4	17.0	1.80	5.6	15.1	2.29*
White boys, 9	80	87.6	15.0		88.0	18.1		−0.4	14.5	
Negro girls, 7	71	92.5	13.9	0.71	81.4	15.0	4.78**	11.2	14.0	4.62**
White girls, 7	83	94.2	15.1		94.2	18.3		0.0	16.0	
Negro girls, 8	61	89.7	13.0	1.06	78.3	18.5	3.48**	11.4	15.1	3.55**
White girls, 8	73	92.4	16.7		90.1	20.8		2.3	14.3	
Negro girls, 9	53	88.2	13.3	0.33	77.6	13.3	3.27**	10.6	12.1	3.78**
White girls, 9	59	87.3	14.2		87.2	17.7		0.1	17.0	

* Significant at the .05 level of confidence.
** Significant at the .01 level of confidence.
[a] Each child's SB IQ minus his CRPM IQ.

The average number of points by which the Negro children's SB IQs exceeded their CRPM IQs was 9.83. For the white children, the average difference was .17 favoring the CRPM. The t value for these differences was 9.26, a significant finding when one considers that, for groups of this size, the t value of 2.58 is significant at the .01 level of confidence. The results, when analyzed by age-color-sex groups, were similar. The Negro children, as a group, obtained significantly higher SB IQs than CRPM IQs. The IQs of the white children were similar on the two tests. Therefore, we accept the null hypothesis that there is no significant difference between verbal and nonverbal intelligence test scores for low socioeconomic status white children, and we fail to accept the null hypothesis that there is no significant difference between verbal and nonverbal intelligence test scores for low socioeconomic status Negro children.

No significant differences were found between Negro boys and girls nor between white boys and girls. There were no differences on the basis of age groups. This is probably accounted for by the decrease in both SB IQs and CRPM IQs with increased age.

The analysis of difference scores revealed a sex difference. The girls averaged 5.56 points lower in their CRPM IQs, and the boys were 2.91 points lower. This difference was significant at the .05 level of confidence. Negro girls tended to have CRPM IQ scores falling lower below their SB IQ scores than did Negro boys, and the white girls tended to have lower CRPM IQs, while the white boys had differences slightly favoring the SB IQs.

Summary

In a population of 789 children of low socioeconomic status, with an age range of seven years to nine years eleven months, no evidence was found to support the hypothesis that social bias in the verbal items of the SB depressed the SB IQ below the nonverbal, nonsocially biased CRPM IQ.

Mean SB IQs were similar for Negro and white children, boys and girls, and for groupings within the three age levels. The CRPM discriminated on the basis of color, with the Negro mean scores in all instances lower than the white mean scores.

No significant difference was found for white children between their SB IQ, which contains many verbal items, and their nonverbal CRPM IQ. However, Negro children's CRPM IQ means were significantly lower than their SB IQ means.

It is suggested that the CRPM cannot be considered a test of intelligence or a measure of *g* but is rather a measure of a specific skill. The present findings suggest that intelligence tests heavily loaded with nonverbal items may discriminate against Negro children.

Received December 10, 1957.

References

1. McNemar, Q. *The revision of the Stanford-Binet scale.* Boston: Houghton Mifflin, 1942.
2. Martin, A. W., & Wiechers, J. E. Raven's Colored Progressive Matrices and the Wechsler Intelligence Scale for Children. *J. consult. Psychol.,* 1954, **18,** 143–144.
3. Ravens, J. C. *Guide to using Progressive Matrices (1947).* London: H. K. Lewis, 1951.
4. *Test Service Bulletin.* New York: Psychological Corp., 1955, No. 49, 6.
5. Vernon, P. E. An application of factorial analysis to the study of test items. *Brit. J. Psychol., Statist. Sect.,* 1950, 3, 1–15.

Journal of Consulting Psychology
Vol. 22, No. 6, 1958

Trail Making Test as a Screening Device for the Detection of Brain Damage

Earl C. Brown,[1] Albert Casey, Ralph I. Fisch,
and Charles Neuringer
VA Center, Wadsworth, Kansas

The Trail Making Test is a relatively unknown instrument, although over the years several investigators have utilized it in making studies. As best as can be determined, the Trail Making Test dates back to a prototype originally known as the Taylor Number Series,[2] a test which required the subject to draw lines connecting a series of numbers from 1 to 50 that were scattered randomly about a rectangular sheet of paper. In 1938, Partington (4) renamed the test "A Test of Distributed Attention" and constructed forms which were modifications of the above. He viewed the test mainly as one that dealt with the speed of motor performance. Further use of the test, however, led Partington to conclude that a successful performance demanded that the subject manifest an integration of previously learned material into a pattern of response that involved shifts in organization, recall, and recognition, as well as motor performance. Later, Partington and Leiter (4) administered the test to an unselected adult male population with the notion of developing a standardized procedure of administration and scoring. When the results were scrutinized, the authors concluded that the test was measuring general mental ability. They demonstrated that the scores on the test were in general agreement with the estimated intelligence of the individuals to whom it was administered. At that time the test was called the Partington Pathways Test.

A follow-up study (4) in 1945 on 256 unselected World War II veterans ranging in age between 19 to 36 years, who had also taken a Stanford-Binet 1937 Revision, revealed a .68 correlation between the Pathways and the latter. The test was then incorporated as a performance subtest in the Army Individual Test of General Ability (7). Again it was rechristened and went under the title of the Trail Making Test. As a part of the Army battery, it underwent an initial validation study conducted on a sample of 465 soldiers, both white and Negro (6). The reliability of the Trail Making was found to be .84 for the white group, with the figure for the Negroes being essentially the same. In final battery form, the test correlated .65 with the Army General Classification Test. This version of Trail Making Test was used in the present study. The War Department released it from classified status and granted the right to reproduce the forms locally. As used in the Army, the goal was primarily one of quick screening of level of intellectual functioning. It was felt that the test would provide an index that was more accurate than one that could be obtained from a group test. However, it was also used on clinical populations and was deemed to have some predictive value in making differentiation between neurotic and schizophrenic groups (8). For civilian purposes, the test has also been put to use by Leiter and Partington (3) as a subtest in an adult performance intelligence scale. Weaver (11) specifically applied the scale at the college level and found that the Trail Making correlated .62 with Q scores of the ACE.

At the same time, Leiter and Partington felt that the test might have some relevance with regard to the evaluation of brain injury. Watson (10) devised a qualitative check list for use in gauging the Pathways' possible value in the evaluation of organics. Armitage (1), employing an altered administration procedure, attempted to use the Trail Making Test as a part of a battery designed specifically for the purpose of evaluating brain injury. He felt that the test would measure a number of functions such as ability to maintain a double relationship, ability to plan foresightfully, ability to make shifts, etc. Using the test, he was able to show, to a highly significant degree, differences between brain damaged *S*s and controls composed of normals and neurotics—the controls making less errors and showing a greater ability to recognize and correct them. Scherer and Winne,[3] however, using the Trail Making over a five-year period as a part of a follow-up study on lobotomized patients, were not able to discover any relationship between Trail Making performance and this type of cerebral deficit. Humphries,[4] using a considerably different administration and scoring procedure, applied the test in group form to organic, schizophrenic, and normal subjects who were matched on the basis of vocabulary scale scores. On this basis, he was able to distinguish both the schizophrenic and organic groups from the normals as well as from each other.

---

[1] Currently Area Chief Psychologist, VA Area Medical Office, Columbus, Ohio.

[2] J. E. Partington. Personal communication. 1956.

Finally, Reitan [5] used the test on a group of highly selected organic patients on which brain damage had been established under intensive and thorough diagnosis. Originally, he studied 84 subjects without evidence of brain damage as a control group and compared them with 200 subjects of known brain damage. The test was able to discriminate well between the groups, with only 16.7% of the controls showing as false positives. In a more rigorous follow-up study, Reitan (5) administered the test individually to 27 patients with brain damage and 27 patients without brain damage. He obtained results which significantly differentiated the two groups ($p < .001$).

It seems clear, then, that Reitan's, Humphries', and Armitage's findings are suggestive of the discriminative and fruitful application of the test to brain injury. However, the approach of Reitan and Armitage involved time-consuming definition of brain injury as well as a projected use of the test for intensive diagnostic purposes. The present study, instead, emphasized its possibilities as a short, simple, and economical screening device for the detection of brain damage. The test was brought into the clinical setting and administered to the patient population as it normally presented itself for diagnostic evaluation, in the hope that such a procedure would shed additional light on the test's value as a practical diagnostic tool.

Procedure

The Trail Making Test was routinely administered to all patients referred to the Clinical Psychology Section by the Psychiatry and Neurology Service, General Medical and Surgical Hospital, and Domiciliary, at the VA Center, Wadsworth, Kansas. The test was administered by members of the Psychology Section from December, 1954, to October, 1956.

Materials

The materials used were two test sheets, Part A and Part B of the Trail Making, and two pencils with erasers. The examiner kept a pencil in his right hand so that he could point to the spots indicated in the directions. He did not draw any of the lines for the examinee but confined himself to pointing. The test was administered in two parts, Part A and Part B. If Part A was failed, Part B was not administered. The time required for completion of each part was noted. Part A required *S*s to draw lines connecting Nos. 1 through 25 in their proper sequence. Part B required following the proper number–letter sequence, 1A-2B3C . . . 12L. The numbers as well as the letters in the case of Part B were enclosed in small circles and scattered randomly over the test sheets. The standard administration procedure as described in the Army Individual Test Manual (7) was used.

---

[3] I. W. Scherer and J. F. Winne. Personal communication. 1956.

[4] C. C. Humphries. Personal communication. 1957.

[5] R. M. Reitan. Personal communication. 1956.

Scoring

For each part passed, the time—in terms of seconds—was noted on the space of the test blank. If the *S* failed, a minus or failure sign was written and the time not recorded.

The scores used for this study were of three types: (*a*) time in seconds to complete Part A, (*b*) time in seconds to complete Part B, and (*c*) space score B, indicating the point in the number–letter sequence at which failure occurred.

Subjects

The diagnostic criterion was the final medical diagnosis assigned by the individual physician in charge of each case.

The subject population for Part A of the test was composed of 39 organic brain-damaged patients, 14 psychotic patients, 29 other NP (neuro-psychiatric) patients with neuroses and character and behavior disorders, and 28 normal patients. For Part B, 16 organic, 7 psychotic, 13 other NP, and 15 normal patients were available. The difference in number of *S*s used for each part of the test was due to the scoring procedure which yielded a time score only if the given part was completed.

Results

The analysis of *S*s' performance in terms of seconds on Part A of the Trail Making Test is summarized in Table 1. In order to meet the assumptions underlying the Analysis of Variance test, it was necessary, first, to apply a reciprocal transformation to the raw scores. Thus, the data were analyzed in the form of reciprocals of seconds × 10,000. Also, inasmuch as the subclass frequencies were unequal, it was necessary to apply special techniques for computing the sums of squares (9, p. 381).

Table 1

Summary of Analysis of Variance for Part A of the Trail Making Test

Source of variation	*df*	Mean square	*F*	*p*
Diagnosis	3	11275.7	5.79	<.01
Age	1	4468.9	2.29	>.05
IQ	1	17104.1	8.78	<.01
Error	94	1947.8		
Interaction Age × IQ	1	4563.9	2.34	>.05
Interaction Age × Diagnosis	3	1677.7	.86	>.05
Interaction IQ × Diagnosis	3	2872.0	1.47	>.05
Interaction Age × IQ × Diagnosis	3	2873.0	1.47	>.05

Table 2

Frequency Distribution of Scores on Part A of the Trail Making Test for the Various Diagnostic Categories

Seconds (Reciprocals ×10,000)	Organic	Psychotic	Other NP	Normal
–40	4			
41–60	5			
61–80	4	3	1	1
81–100	2		1	
101–120	4	1	1	1
121–140	7	5	3	1
141–160	4	1	1	4
161–180	2	1	4	4
181–200	1	3		3
201–220				2
221–240	1		3	6
241–260	1		1	1
261–280	1		4	1
281–300	2		2	2
301–320				
321–340	1		4	
341–360			1	
361–380				
381–400			1	
401–			2	2

Significant differences were found to exist between the diagnostic categories and between the high and low IQ groups, the high IQ group taking significantly less time to complete this part of the test.

The means of the diagnostic groups were: organic 127.4; psychotic 130.5; other NP 248.8; normal 217.6. To determine the locus of the significant differences obtained between the diagnostic groups, the means were treated by Duncan's multiple range test (2). Because of unequal *N*s, calculations were based on the smallest group with an *N* of 14. According to Duncan's test, the minimum difference between means for significance at the 5% level is no greater than 35.99. By this criterion, the differences between all means are significant with two exceptions, occurring between the organic and psychotic groups on the one hand and the other NP and normal groups on the other.

Table 2 represents the frequency distribution of the seconds scores on Part A.

Obviously, the large amount of overlap between the groups, especially the organic and psychotic groups, rules out the possibility of establishing a meaningful cutting score that would distinguish clearly between the organic and the other diagnostic groups of *S*s in the present population. Treatment of *S*'s scores in terms of seconds on Part B of the test is outlined in Table 3.

Due to the relatively small *N*s present in this analysis, it was necessary to pool and consider as a unit all nonorganic groups of *S*s, i.e., the psychotic, other NP, and normal groups. As is seen from inspection of Table 3, the only comparisons yielding significant differences were between the low and high IQ groups and between the two age groups. Again, it is clear from the frequency distribution of these scores (Table 4) that the establishment of a practical cutting score on the basis of the present data is not possible.

Finally, in order to compare more clearly the present results with Reitan's (5) findings, a frequency distribution comprised of coded scores for Part A plus Part B is given in Table 5. Such scores are obtained by application of the Army Test Manual scoring procedure (7)—the same procedure that Reitan followed in his studies. *S*s failing Part B were included in this table, in which case a score of zero was given for Part B.

Table 3

Summary of Analysis of Variance for Part B of the Trail Making Test

Source of variation	*df*	Mean square	*F*	*p*
Diagnosis	1	360.4	.79	>.05
Age	1	3196.0	7.0	<.01
IQ	1	3164.1	6.9	<.01
Error	43	453.5		
Interaction Age × IQ	1	94.6	.21	>.05
Interaction Age × Diagnosis	1	2.3	.005	>.05
Interaction IQ × Diagnosis	1	6.7	.01	>.05
Interaction Diagnosis × Age × IQ	1	1.1	.002	>.05

Table 4

Frequency Distribution of Scores on Part B of the Trail Making Test for the Various Diagnostic Categories

Seconds (Reciprocals ×10,000)	Organic	Psychotic	Other NP	Normal
–20	1			
21–30	1			1
31–40		2		
41–50				
51–60		1	2	3
61–70	4	1		1
71–80	5	2		1
81–90	2		2	1
91–100				3
101–110	1		1	
111–120			4	2
121–130				1
131–140	1	1		
141–150				1
151–160			1	
161–	1		3	1

Reitan used a cutting score falling between 12 and 13. Eighty-three per cent of his control group, composed of normal and other NP *S*s, fell above this score in contrast to only 17% of the brain-damaged group tested. The present data, however, present quite a different picture. Inspection of Table 6 indicates an exorbitant number of false positives associated with the use of this cutting score. It is clear from the frequency distributions of Part A and Part B scores (Tables 2 and 4), considered separately or together, that this circumstance is due primarily to the inclusion, in the present data, of the scores of the psychotic group which closely parallel the scores of the organic group.

An additional measure, space scores, was investigated for the purpose of testing the hypothesis that, among the *S*s failing Part B, the organic *S*s would make errors earlier in the number–letter sequence than would the nonorganic *S*s. A space score simply represented the last number correctly reached be-

Table 5

Frequency Distribution of Coded Scores on Part A plus Part B of the Trail Making Test

Scores[a]	Organic	Nonorganic
20		
19		3
18		1
17	1	1
16		
15	1	1
14		5
13	1	1
12		2
11		4
10	2	8
9	2	10
8	1	5
7	1	5
6	6	9
5	3	3
4	5	5
3		4
2	4	2
1	12	2
	N = 39	*N* = 71

[a] Reitan's cutting score is between 12 and 13.

fore an error occurred. Unfortunately, due to the short range within which these scores may fluctuate, they could neither be normalized nor ranked properly for accurate statistical evaluation. Nevertheless, inspection of the frequency distribution of these scores indicated that this measure offered little if any advantage over the previous measures considered.

Summary and Conclusions

The present study was undertaken in the hope of finding in the Trail Making Test a shorter and more economical method of screening for organic brain damage. A number of earlier studies, especially those of Reitan, Humphries, and Armitage, strongly suggested the possibility of such an application for this test. The approach, although less rigorous than that used in the previously mentioned studies, had the advantage of more closely approximating the conditions under which the problem of practical application of the test could be subjected to direct inquiry, i.e., conditions of clinical practice. In this respect, the findings were essentially negative.

1. Time scores for Part A of the Test were found to significantly differentiate between the high and low IQ groups and as well between the organics and psychotics as one grouping and the normals and other NP patients as another.

2. Part B was found to differentiate only between the high and low IQ groups and between the high and low age groups.

3. Frequency distributions yielded by both measures, considered separately or together in the form of coded scores, did not offer the possibility of a cutting score that would have discriminative diagnostic value with respect to the present population.

4. An additional measure considered, space scores, was likewise found to lack discriminative value.

Inasmuch as the diagnostic problem frequently encountered in clinical practice is one of distinguishing between psychosis and brain damage, the present findings clearly suggest a verdict of "no value" for the Trail Making Test in this context. Suggested, further, is that any application of the test to a population of the type presently considered must take into account factors of age and IQ.

It is interesting to note that, contrary to the experimenters' expectations, Part A seemed more sensitive to diagnostic categories than Part B. Part A certainly requires less "distributed attention." It seems rather to more closely approximate a simple speed-of-reaction task.

Finally, it is suggested that if the merit and possible applications of the Trail Making Test are to be determined, certain areas of investigation should be pursued. Among the most outstanding is the possibility for developing methods of analyzing errors in a more qualitative fashion. Also, relationships between age, IQ, method of establishing diagnosis, and Trail Making scores need further understanding.

Received November 22, 1957.

References

1. Armitage, S. G. An analysis of certain psychological tests used in evaluation of brain injury. *Psychol. Monogr.*, 1946, 60, No. 1 (Whole No. 277).

2. Duncan, D. B. Multiple range and multiple *F* tests. *Biometrics,* 1955, **11**, 1–42.
3. Leiter, R. G., & Partington, J. E. *Examination manual for the Leiter-Partington Adult Performance Scale.* Washington: Psychol. Serv. Center Press, 1950.
4. Partington, J. E., & Leiter, R. G. Partington Pathways Test. *Psychol. Serv. Center Bull.,* 1949, **1**, 9–20.
5. Reitan, R. M. The relation of the Trail Making Test to organic brain damage. *J. consult. Psychol.,* 1955, **19**, 393–394.
6. U. S. War Department, Adjutant General's Office, Classification and Replacement Branch, Staff, Personnel Research Section. The new Army Individual Test of General Mental Ability. *Psychol. Bull.,* 1944, **41**, 532–538.
7. U. S. War Department, Adjutant General's Office. *Manual of the Army Individual Test.* Washington: Author, 1944.
8. U. S. War Department, Adjutant General's Office. *The relationship of Army Individual Test subscores and other mental ability tests to diagnosis of mental disorder.* (P.R.S. Report No. 724.) Washington: Author, 1946.
9. Walker, H. M., & Lev, J. *Statistical inference.* New York: Holt, 1953.
10. Watson, N. A qualitative check list for clinical use of Partington's Pathways Test in evaluating brain injury. *Psychol. Serv. Center J.,* 1949, **1**, 49–51.
11. Weaver, H. B. The Leiter-Partington Adult Performance Scale at the college level. *Psychol. Serv. Center J.,* 1950, **2**, 182–188.

Journal of Consulting Psychology
Vol. 22, No. 6, 1958

Characteristics of Volunteers and Nonvolunteers in Psychological Experimentation[1]

R. M. Martin and F. L. Marcuse
State College of Washington

A social psychology or generalized conclusion based on experimental investigations of freshmen or sophomores in college is highly debatable. Use of volunteers from such freshmen or sophomore classes is even more debatable for it may represent a selection within a selection. In experimental work it is also a wise precaution to know as many characteristics as possible of the group being investigated.

Studies have been made of the scores of volunteers on visuo-motor learning (4), the type of person requesting volunteers (5), marital adjustment of volunteers (15), as well as the nature of the volunteering act itself (2, 9, 10, 11). Maslow and Sakoda (8) found volunteers to be higher in a scale of self-esteem than nonvolunteers and, as a result of their findings, concluded that the Kinsey data on sex (6, 7) should allow for volunteer error. Siegman (12), however, using a number of personality variables, as well as the self-esteem scale used by Maslow, found no differences between volunteers and nonvolunteers for a Kinsey-type interview.

The present study proposed to examine the personality characteristics of volunteers and nonvolunteers for different experimental situations. Reliability of the volunteering act was also to be examined.

Procedure

Approximately 400 introductory psychology students at the State College of Washington were used during the different phases of this study. Four steps, each separated by approximately one week, were involved. They were as follows: Step 1 involved the administering of a battery of tests. Such tests consisted of the Taylor Manifest Anxiety Scale (13), the Levinson E (ethnocentrism) Scale (1), and the administration of the Bernreuter Personality Inventory of which four scales were used (3). Scores made on the American Council on Education Psychological Examination (14) were also obtained. Step 2 requested volunteers. One of the experimenters (Martin) met with each of 12 sections and presented a volunteer request sheet for one of four experimental situations. The situations were as follows: One asked for volunteers for an experiment dealing with learning, a second for personality, a third for attitude to sex, and the final one asked for volunteers for hypnosis. Three sections were used for each specific experimental situation. The written request sheet asking for volunteers was made as traditional as possible and is reproduced here verbatim:

> The Department of Psychology would like to know if you can be a volunteer for an experiment dealing with [at this point either learning, personality, attitude to sex, or hypnosis were inserted]. This experiment will take approximately two hours of your time. Volunteering or not volunteering will have no effect on your grade in this course. On the sheet below please sign your name if you can be a volunteer. For those of you who indicate you can be volunteers, contact will be made at a later date to inform you of the different times available and of the place.

Step 3 was designed to check the reliability of volunteering. The procedure decided on was as follows. A week after the original request, the same experimenter again met the nonvolunteers. Table 3 shows that reliability of volunteering for the different experimental situations varied from .67 to .97. Reliability for all four groups considered as a unit was .86.

---

[1] The scope of this investigation was made possible by Grant 3220-110 of the State College of Washington. This study was done in partial fulfillment of the requirements for the degree of master of science in psychology.

Discussion

When volunteers and nonvolunteers are each considered as a unit for the four experimental situations, they show no difference in six of the seven personality variables. The one exception to this was the finding that volunteers were significantly more intelligent (as defined by ACE score) than nonvolunteers. The reason for this may be the fact that the more intelligent person is more inquisitive and that volunteering might reflect this. The data indicated that female and male volunteers for personality have a suggestive higher manifest anxiety level than nonvolunteers. It is possible that there is something implicit in the investigation of personality that attracts individuals that are more anxious and who feel the need of personal aid or information that such research might seem to offer. In the request for volunteers for attitudes on sex, no significant differences were found in any comparison. This finding has direct bearing on an issue that has been of some concern. It was pointed out that the publication of the Kinsey reports (6, 7) brought up the question of volunteer error which Maslow and Sakoda (8) believed would bias its results. Siegman (12), on the other hand, found no differences between volunteers and nonvolunteers on such variables as anxiety, rigidity, defensiveness, nor on the Maslow scale of self-esteem, and concluded that the problem of volunteer error was not present in the Kinsey studies. The findings of the present study would tend to support the view of Siegman (although the reliability of the volunteering response for this situation [.67] may cloud the finding to some extent) and is not too surprising when one considers the myriad of ways employed by Kinsey in obtaining volunteers. Volunteers for hypnosis as a group were significantly less ethnocentric and had a suggestively higher intelligence score than nonvolunteers. This, and the very fact that no significant differences were found in any comparison between volunteers and nonvolunteers for hypnosis in self-sufficiency, dominance–submission, anxiety, and introversion–extroversion, was considered important in that popular belief suggests the contrary.

While the reliability of the volunteering response in this study was considered satisfactory, its definition should be noted. It is defined in terms of verbal indication of willingness to volunteer, and it may be argued that there is a gap between what individuals say they will do and their actual behavior. It should be realized that the data do not indicate that those who volunteer for one experimental situation would volunteer for another situation. The data only indicate the reliability of volunteering for a given situation and show that regardless of the experimental situation "request" reliability was satisfactory.

It is interesting to note that Wilcox and Faw (16), using different tests and describing the psychological traits of susceptible and nonsusceptible hypnotic *Ss* (individuals who actually participated in an experiment rather than volunteers and nonvolunteers), found similar results to those reported here.

Himmelstein (5) has indicated that the appearance of the individual requesting volunteers as well as the nature of the request (and often in conjunction with it) might be important in determining the magnitude of the response. This factor in the present study was controlled by having the same experimenter, a presentable, young (24 years) white male make all the requests. It would be interesting to conjecture what the magnitude of male–female response might have been to requests for volunteers for an experiment dealing with sex or hypnosis if the experimenter had been a presentable, young white female!

Summary and Conclusion

This study concerned the personality characteristics of volunteers and nonvolunteers for different types of psychological experiments. Four specific experimental situations (learning, personality, attitude to sex, and hypnosis) requesting volunteers were used. Three sections of introductory psychology were used for each specific request situation. The total number of *Ss* used was approximately 400. The request sheets asking for volunteers were traditional. Comparisons were made of differences between volunteers and nonvolunteers for each and for all of the four experimental situations on the seven personality variables. Reliability of the volunteering act was checked.

Significant differences were found on two of the seven personality variables between volunteers and nonvolunteers for one of the four specific request situations (hypnosis). Suggestive differences were also indicated. Reliability for the different volunteering situations was found to vary from .67 to .97.

The general conclusion of this investigation was that there were personality differences between volunteers and nonvolunteers associated with different types of volunteering situations and that generalizations made from biased samples can obviously be misleading. The general practice of using volunteers probably owes its wide application to matters of expediency. However, it does not seem that convenience should substitute for sound experimental procedure.

Received December 2, 1957.

References

1. Adorno, T. W., Frenkel-Brunswik, Else, Levinson, D. J., & Sanford, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. Blake, R. R., Berkowitz, H., Bellamy, R. Q., & Mouton, J. S. Volunteering an avoidance act. *J. abnorm. soc. Psychol.*, 1956, **53**, 154–156.
3. Bernreuter, R. G. *The personality inventory*. Stanford, Calif.: Stanford Univer. Press, 1935.
4. Brower, D. The role of incentive in psychological research. *J. gen. Psychol.*, 1948, **39**, 145–147.
5. Himelstein, P. Taylor scale characteristics of volunteers and nonvolunteers for psychological experiments. *J. abnorm. soc. Psychol.*, 1956, **52**, 138–139.
6. Kinsey, A. C., Pomeroy, W. B., & Martin, C. E. *Sexual behavior in the human male*. Philadelphia: Saunders, 1948.
7. Kinsey, A. C., Pomeroy, W. B., Martin, C. E., & Gebhard, P. H. *Sexual behavior in the human female*. Philadelphia: Saunders, 1953.
8. Maslow, A. H., & Sakoda, J. M. Volunteer error in the Kinsey study. *J. abnorm. soc. Psychol.*, 1952, **47**, 259–262.
9. Rosen, E. Differences between volunteers and nonvolunteers for psychological studies. *J. appl. Psychol.*, 1951, **35**, 185–193.
10. Rosenbaum, M. E. The effect of stimulus background factors on the volunteering response. *J. abnorm. soc. Psychol.*, 1956, **53**, 118–121.
11. Rosenbaum, M. E., & Blake, R. R. Volunteering as a function of field structure. *J. abnorm. soc. Psychol.*, 1955, **50**, 193–196.
12. Siegman, A. Responses to a personality questionnaire by volunteers and nonvolunteers to a Kinsey interview. *J. abnorm. soc. Psychol.*, 1956, **52**, 280–281.
13. Taylor, Janet. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
14. Thurstone, Thelma G., & Thurstone, L. L. *American council on education psychological examination for college freshmen*. New York: Educational Testing Service, 1953.
15. Wallin, P. Volunteer subjects as a source of sampling bias. *Amer. J. Sociol.*, 1949, **54**, 539–544.
16. Wilcox, W. W., & Faw, V. Psychological traits of susceptible and unsusceptible hypnotic subjects. Paper read at Western Psychological Association, Eugene, Oregon, May, 1957.

Journal of Consulting Psychology
Vol. 22, No. 6, 1958

"Not Alike" Responses in Wechsler's Similarities Subtest[1]

Murray Levine
Devereux Foundation Institute of Research & Training, Devon, Pennsylvania

Wiener (2) recently found that the "not alike" error in the Similarities subtest increased in frequency as a function of instructions designed to create suspicion. He also found that characteristically suspicious *S*s, as independently measured by a scale of distrust, had greater intellectual deficit than more trusting *S*s. One could partially confirm his findings if one could show that *S*s who gave the "not alike" response spontaneously had a lower mean IQ than did *S*s who did not make that particular error. If distrust or suspiciousness pervaded the test performance of these *S*s, the mean difference in IQ would be greater than could be accounted for by the lost IQ points due to the error itself.

The 400 *S*s are veterans with a wide variety of psychiatric diagnoses who had been referred for psychological testing in an outpatient setting.

Two hundred "not alike" *S*s were selected from our files on the basis of the appearance of one or more "not alike" responses to Similarities test items of the Wechsler-Bellevue Form I. The criteria for the "not alike" response were quite stringent. Control *S*s were selected by taking the next case in alphabetical order to the man who made the error. Unrecorded errors that were made would act against the hypothesis.

There is a highly significant difference in IQ in favor of the control *S*s. The mean IQ for the control group is 109.5 ($SD = 15.03$), while the mean for the not alike group is 100.8 ($SD = 14.39$). The mean difference between groups yields a t value of 5.92, which is significant well beyond the .001 level of confidence. This difference in IQ is considerably greater than can be accounted for only on the basis of the "not alike" error. Loss due to "not alike" responses alone amounts to a *maximum* of three IQ points. The obtained loss is 8.7 IQ points. Even if we subtract 3 IQ points from the obtained mean difference, using the same standard error of the difference, the t value would still be significant beyond the .01 level of confidence.

The present finding, showing that *S*s who make the "not alike" error have a lower mean IQ than *S*s who do not, is in partial confirmation of Wiener's previous work with the attitude of distrust and its relationship to intelligence. Suggestively, the attitude of distrust serves to reduce effectiveness in intelligence test performance.

Because of the substantial differences in IQ found in this and a previous study of a qualitative error (1), it is our impression that further detailed analysis of intelligence test performance in the light of current concepts of ego functioning will point the way to integrating theories of intelligence within a more general theory of personality.

Brief Report.
Received May 15, 1958.

References

1. Levine, M., Glass, H., & Meltzoff, J. The inhibition process, Rorschach human movement responses, and intelligence. *J. consult. Psychol.*, 1957, **21**, 41–45.
2. Wiener, G. The effect of distrust on some aspects of intelligence. *J. consult. Psychol.*, 1957, **21**, 127–130.

---

[1] An extended report of this study may be obtained without charge from Murray Levine, Ph.D., Devereux Foundation Institute for Research and Training, Devon, Pennsylvania, or for a fee from the American Documentation Institute. Order Document No. 5693, remitting $1.25 for microfilm or $1.25 for photocopies.

Journal of Consulting Psychology
Vol. 22, No. 6, 1958

Personal Adjustment, Assumed Similarity to Parents, and Inferred Parental-Evaluations of the Self

Melvin Manis[1]
Ann Arbor VA Hospital, Michigan

Many clinicians have observed that the maladjusted individual tends to feel isolated and different from the important figures in his environment. One aim of the present study was to provide a quantitative test of this hypothesis. More specifically, in view of the importance of parental relationships, it was predicted that the maladjusted individual sees himself as being less like his parents than does the person who is relatively free of emotional problems.

Two cases cited by Mowrer (14, pp. 531–535) led to the second hypothesis of the study. Using the semantic differential (15), he reported that a male patient at the outset of therapy rated the concepts "me" and "mother" as being more similar than the concepts "me" and "father"; at the termination of therapy this pattern was reversed (i.e., "me" was rated more similar to "father" than to "mother"). Similarly, a female patient at first saw more similarity between the concepts "me" and "father" than between the concepts "me" and "mother," with this pattern showing a reversal at the termination of therapy. These observations suggest the hypothesis that people with emotional disturbances see themselves as being more similar to their parent of the opposite sex than to their parent of the same sex, while people who are relatively well-adjusted see themselves as being more like their same-sexed parent.

The third hypothesis of the study was that the maladjusted individual perceives his parents as differing in their evaluations of him. There are two lines of reasoning which lead to this prediction. Both of these formulations are related to Brownfain's finding (2) that people with unstable self concepts, i.e., those who show a marked discrepancy between their positive (optimistic) and negative (pessimistic) self-descriptions, tend to be poorly adjusted. From one viewpoint, perceived parental disagreement may be regarded as a source of anxiety which would be particularly potent in contributing to the instability of the neurotic's self concept. On the other hand, the direction of causality may be reversed. That is, since perceived parental evaluations are probably colored by autistic factors, the maladjusted person may project his conflicting beliefs about himself onto his parents and conclude that their evaluations of him differ widely.

Method

Two groups of students, one well-adjusted and one poorly adjusted, were selected from an introductory psychology course; there were 15 males and 15 females in each group. Personal adjustment was defined in terms of the *S*'s performance on the *D*, *Pt*, *Hs*, *Hy*, *L*, and *K* scales of the MMPI (9). These scales were combined into an over-all measure of adjustment by averaging the *S*s' *T* scores on the four clinical scales (including correction factors). In selecting *S*s, males and females were matched on the basis of this adjustment index. The average *T* score for the adjusted *S*s was 48.3, the average for the maladjusted *S*s was 64.1. *S*s who had a raw score above 5 on the *L* scale, or who left out 17 or more items, were not included in these groups. In addition, no *S* was included in the adjusted group if he had a T score of 60 or above on any one of the clinical scales.

---

[1] This study was conducted while the author held a Public Health Service career-teacher grant at the University of Pittsburgh.

The *S*s described their "real" selves, their "ideal" selves, and each of their parents on a series of 24 bipolar rating scales; the *S*s also described themselves as they thought they were seen by each of their parents. The rating scales, which were presented in a seven-interval format, were derived from Cattell's (3) factor analysis of the Allport-Odbert adjective trait list (1). Cattell's analysis resulted in 12 relatively independent factors. In the present study, the eight factors which accounted for the most variance were represented by three scales each. The items may therefore be assumed to sample most of the important dimensions that English-speaking people use in describing themselves and others. Some typical scales were frank–secretive, immature–mature, impatient–patient, and intelligent–unintelligent.

The similarity between the *S*s' self-descriptions and their descriptions of their parents was quantified in terms of the D statistic (6, 16). This variable will hereafter be referred to as assumed similarity (AS). The similarity between each *S*'s inferred father-evaluation and his inferred mother-evaluation was also quantified using the D score.

Results and Discussion

Personal Adjustment and Assumed Similarity to Parents

The experimental design for the assumed similarity data consisted of a 2 × 2 factorial design (adjustment × sex) in which two scores were available for each *S*: AS to the same-sexed parent and AS to the opposite-sexed parent. These data were analyzed in accordance with the Type III mixed design discussed by Lindquist[2] (13, pp. 231–234). The results of this analysis are presented in Table 1.

Table 1

Analysis of Variance for Assumed Similarity Scores

Source	*df*	Mean square	*F*	*p**
Between Subjects	59			
S's adjustment	1	91.98	11.04	<.01
S's sex	1	79.42	9.53	<.01
S's adjustment × *S*'s sex	1	20.19	2.42	
Error (b)	56	8.33		
Within Subjects	60			
Parent's sex[a]	1	1.14		
Parent's sex × *S*'s adjustment	1	4.06		
Parent's sex × *S*'s sex	1	1.54		
Parent's sex × *S*'s adj. × *S*'s sex	1	.01		
Error (w)	56	4.26		
Total	119			

[a] AS to same-sexed vs. opposite-sexed parent.
* Two-tailed test.

In this analysis, each of the mean squares in the "Between Subjects" part of the table was evaluated against Error (b), while the mean squares from the "Within Subjects" portion were evaluated against Error (w). Significant *F*s ($p < .01$) were obtained for variance due to adjustment and for variance due to the *S*'s sex. That is, when the two assumed similarity scores for each S were totalled, the well-adjusted groups (both male and female) showed more assumed similarity to their parents on this composite measure than did the poorly adjusted groups. In the comparison by sex, males showed more similarity to their parents on the composite score than did females.

The results of this analysis do not show any significant differences between assumed similarity to the same-sexed and the opposite-sexed parent. This finding was obtained at both levels of adjustment, as may be seen by the nonsignificant interaction terms in the "Within Subjects" portion of Table 1.

These results support the hypothesis that well-adjusted people show more assumed similarity to their parents than do those with emotional disturbances. There was, however, a possibility of artifact: if most of the parent-descriptions had been highly favorable, i.e., similar to the *S*s' ideal selves, well-adjusted *S*s, having high self-esteem, would of necessity show high AS.[3] Careful examination of the data did not, however, support the basic assumptions of this view. Instead of being consistently favorable, the parent-descriptions varied widely in favorability and were, on the average, slightly *less* favorable than the *S*s' self-descriptions.[4]

---

[2] Inspection of the data suggested that they did not depart markedly from normality. In addition, application of Bartlett's test yielded a nonsignificant result, indicating that the variance within groups was relatively homogeneous.

The obtained relationship between personal adjustment and AS is in general accord with previous findings (4, 10, 11, 12, 17). These results indicate that the maladjusted person typically feels that he is quite different from his parents. In addition, since previous research by Jourard (11) and by Fiedler, Warrington, and Blaisdell (7) has shown that high AS may reflect liking for the "other person," these results suggest that the maladjusted person feels less warmly toward his parents than does the well-adjusted person.

The hypothesis that well-adjusted *S*s assume more similarity to their parent of the same sex than to their parent of the opposite sex was not supported; nor was there any evidence that poorly adjusted *S*s assume greater similarity to their opposite-sexed parent. AS to the same-sexed parent was about the same as AS to the opposite-sexed parent for both the well-adjusted and the poorly adjusted *S*s. It is conceivable that different results would have been obtained if the items had been specifically selected to represent areas in which quite different behavior is required of men and women. In this study, however, no such attempt was made; instead, the items were selected to provide a more global coverage of the *S*s' self concepts and their perceptions of their parents.

The fact that women showed less AS to their parents than did men was unexpected. It may be that this finding is related to differences between the typical parent-son and parent-daughter relationship. If, as Fiedler (8) hypothesizes, high assumed similarity can be interpreted as a sign of interpersonal warmth and closeness, these data suggest that women feel more distant from their parents than do men. It should be noted that in spite of the obtained sex difference, the relationship between AS and personal adjustment held for both men and women.

Table 2

Analysis of Variance for Perceived Parental Conflict

Source	*df*	Mean square	*F*	*p**
Between groups	3	19.82	3.42	<.05
Adjustment	1	35.27	6.08	<.05
Sex	1	15.33	2.64	
Adjustment × sex	1	8.86	1.53	
Within groups	52	5.80		
Total	55			

* Two-tailed test.

Personal Adjustment and Perceived Conflict in Parental Evaluations of the Self

To test the hypothesis that the maladjusted person perceives his parents as differing in their evaluations of him, the discrepancies between the inferred parent-evaluations were analyzed in a 2 × 2 analysis of variance (adjustment × sex).[5, 6] The results of this analysis are presented in Table 2. The analysis indicates that there was a significant difference ($p < .05$) between the adjusted and the maladjusted *S*s in the hypothesized direction—the adjusted *S*s showed smaller discrepancies between their two inferred parent-evaluations than did the maladjusted *S*s.

The *D* scores, which constituted the measure of perceived parental conflict, can be divided into two components. Part of the *D* is due to differences between the *elevations* of the inferred parent-evaluations. Thus, an *S* may obtain a high *D* score because he thinks that one parent respects him greatly while the other holds him in low esteem. Differences between the *contents* of the inferred evaluations constitute the second component. An *S* who thinks that both parents esteem him equally might still show a large *D* because of his belief that his father would rate him high on Traits *a*, *b*, *c*, and low on Traits *d*, *e*, *f*, while his mother would do the opposite.

---

[3] Child (5) has noted a similar artifact possibility in reviewing a study by Sopchak (17).

[4] The favorability of each parent-description was quantified by comparing it with the *S*'s ideal self, the *D* statistic (6, 16) being the measure of similarity. The favorability of self-descriptions was assessed in the same fashion.

[5] Incomplete data made it necessary to eliminate one *S* from each of the originally selected groups. As a result, this analysis included data from 28 adjusted and 28 maladjusted *S*s; there were 14 men and 14 women within each adjustment category.

[6] Once again, the data were consistent with the assumptions underlying the analysis of variance technique (see Footnote 2).

The adjusted *S*s felt that they were more highly esteemed by their parents than did the maladjusted *S*s. As a result, the adjusted *S*s might *necessarily* have shown smaller "elevation" differences, since, if the less favorable of their inferred parent-evaluations had been sufficiently high, the more favorable one could not have been much higher. The adjusted group's smaller *D* scores might therefore have been due to a restriction on the size of their "elevation" differences.

To rule out this possibility it was necessary to demonstrate that "elevation" differences were of equal magnitude for the adjusted and the maladjusted *S*s. A new set of discrepancy scores solely reflective of "elevation" differences was therefore computed. These scores were obtained by (*a*) calculating the *D* between each parent-evaluation and the *S*'s ideal self and (*b*) taking the absolute difference between these *D*s. An analysis of variance similar to the one reported in Table 2 was then performed on the resultant measures.

The results of this analysis indicated that the adjusted and the maladjusted *S*s did *not* differ in terms of their "elevation" difference scores; in fact, the variance due to adjustment was *smaller* than the within groups variance. This negative finding means that the results of Table 2 were not due to a restriction on the range of "elevation" differences which could be obtained by the adjusted *S*s.

The negative results of this second analysis also mean that the maladjusted person does not have any marked feeling that one parent regards him more favorably than the other. His relatively large *D* score is apparently due primarily to *content* differences. That is, he may feel that while both parents hold him in relatively low esteem, they do so for different reasons. As a hypothetical example, a maladjusted *S* might believe that one parent thinks of him as being secretive and short-tempered, while the other does not accept this characterization, but does regard him as being undependable and immature.

While this study has related personal adjustment to *perceived* conflict between parental evaluations of the self, the data do not enable us to assess the accuracy of the *S*s' perceptions. Subsequent research might profitably attack the question of whether there is a *real* conflict between the parental evaluations of the maladjusted person.

Summary

Two groups of *S*s, one well-adjusted and one poorly adjusted, were selected from an introductory psychology course. The *S*s described themselves and each of their parents on 24 bipolar rating scales; they also described themselves as they thought they were seen by each parent. The *D* statistic was used to quantify (*a*) the similarity between the *S*s' self-descriptions and their descriptions of their parents and (*b*) the similarity between each *S*'s two inferred parent-evaluations.

The results supported the hypothesis that well-adjusted people perceive themselves as being more like their parents than do those who are poorly adjusted. There was no support, however, for the hypothesis that well-adjusted people perceive themselves as being more similar to their parent of the same sex than to their parent of the opposite sex; nor was there any evidence that poorly adjusted people see themselves as being more like their opposite-sexed parent. An unexpected finding was that women perceive themselves as being less like their parents than do men. According to one viewpoint (8), these results suggest that women feel more distant from their parents than do men.

As hypothesized, the data also indicated that the maladjusted *S*s felt that there was a greater disparity between their parents' evaluations of them than did the adjusted *S*s. Further analysis revealed that this was not due to any marked feeling among the maladjusted *S*s that one parent regarded them more favorably than did the other. Instead, their large *D*s were primarily due to content differences. For example, a maladjusted *S* might feel that his father perceived him as being secretive and short-tempered, while his mother did not accept this characterization, but in-

stead regarded him as being undependable and immature.

Received December 6, 1957.

References

1. Allport, G. W., & Odbert, H. S. Trait names: A psycho-lexical study. *Psychol. Monogr.,* 1936, 47, No. 1 (Whole No. 211).
2. Brownfain, J. J. Stability of the self concept as a dimension of personality. *J. abnorm. soc. Psychol.,* 1952, 47, 597–606.
3. Cattell, R. B. *Personality: A systematic, theoretical, and factual study.* New York: McGraw-Hill, 1950.
4. Cava, F. L., & Raush, H. Identification and the adolescent boy's perception of his father. *J. abnorm. soc. Psychol.,* 1952, 47, 855–856.
5. Child, I. L. Personality. *Annu. Rev. Psychol.* 1954, 5, 149–170.
6. Cronbach, L. J., & Gleser, Goldine C. Assessing similarity between profiles. *Psychol. Bull.,* 1953, 50, 456–473.
7. Fiedler, F. E., Warrington, E. G., & Blaisdell, F. J. Unconscious attitudes as correlates of sociometric choice in a social group. *J. abnorm. soc. Psychol.,* 1952, 47, 790–796.
8. Fiedler, F. E. The psychological-distance dimension in interpersonal relations. *J. Pers.,* 1953, 22, 142–150.
9. Hathaway, S. R., & McKinley, J. C. *The Minnesota Multiphasic Personality Inventory.* New York: Psychological Corp., 1951.
10. Helper, M. M. Learning theory and the self concept. *J. abnorm. soc. Psychol.,* 1955, 51, 184–194.
11. Jourard, S. M. Identification, parent cathexis, and self esteem. *J. consult. Psychol.,* 1957, 21, 375–380.
12. Lazowick, L. On the nature of identification. *J. abnorm. soc. Psychol.,* 1955, 51, 175–183.
13. Lindquist, B. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
14. Mowrer, O. H. *Psychotherapy theory and research.* New York: Ronald, 1953.
15. Osgood, C. E. The nature and measurement of meaning. *Psychol. Bull.,* 1952, 49, 197–237.
16. Osgood, C. E., & Suci, G. H. A measure of relation determined by both mean difference and profile information. *Psychol. Bull.,* 1952, 49, 251–262.
17. Sopchak, A. Parental "identification" and "disorder tendency." *J. abnorm. soc. Psychol.,* 1952, 47, 159–165.

Journal of Consulting Psychology
Vol. 22, No. 6, 1958

A Comparison of the Edwards PPS Variables with Some Aspects of the TAT[1, 2]

Tom Dilworth IV[3]
Southern Methodist University

The Edwards Personal Preference Schedule (PPS) is based on 15 personality needs which are also frequently used in the evaluation of Thematic Apperception Test (TAT) protocols. It would appear, therefore, that the two tests seek to elicit comparable information with regard to needs. The present study investigated the correlation between the 15 PPS variables and the protocols derived from 10 selected pictures of the TAT. Of particular interest was assessment of the value of the PPS as a reasonable substitute for the TAT in clinical situations which may make such replacement necessary or desirable. A significantly positive correlation could indicate that the PPS (which usually requires less time and skill for interpretation than the TAT) might be used with reasonable confidence as an economical and valid substitute for the TAT in cases where time or skill of the interpreter is a determining factor.

Five qualified clinical psychologists[4] independently evaluated the 10-story TAT protocols of 20 college males with respect to the relative strengths of the 15 needs reflected by the PPS also. The null hypothesis, as follows, was confirmed: That there is no significantly positive correlation between the relative strengths of the 15 personality needs represented intra-individually by the PPS and the relative strengths of these same needs as indicated intra-individually by the TAT.

The mean Spearman rank-difference correlation between the relative strengths of TAT protocol needs and the relative strengths of PPS needs was .15. This correlation would not indicate the feasibility of using the PPS as an economical substitute for the TAT. Only 2 of the 20 mean coefficients from which the over-all mean of .15 was derived were significantly positive ($p = .01$). Furthermore, in the case of one judge only was there significant agreement between the five highest-ranked TAT needs and the five highest-ranked PPS needs for each subject ($p = .10$).

Reasonably high interjudge agreement on TAT need ranks was reflected by a mean Kendall coefficient of concordance of .40 ($p = .02$). Additional support for the reliability of the TAT judgments was provided by the fact that there was homogeneity among the judges in their ranking of TAT protocol needs.

Brief Report.
Received June 23, 1958.

---

[1] An extended report of this study may be obtained without charge from Tom Dilworth IV, P.O. Box 1243, Lexington, Kentucky, or for a fee from the American Documentation Institute. Order Document No. 5742, remitting $1.75 for microfilm or $2.50 for photocopies.

[2] This report is based on a thesis submitted to the graduate school of Southern Methodist University in partial fulfillment of the requirements for the M.A. degree. The author is indebted to Robert E. Stoltz for advice and assistance during the course of the study.

[3] Now at the University of Kentucky.

[4] The psychologists were William R. Garretson, John H. Gladfelter, Joseph H. Siegel, Aubrey E. Wilkinson, and Harold R. Winer.

Journal of Consulting Psychology
Vol. 22, No. 6, 1958

The PALS Tests: A Technique for Children to Evaluate Both Parents[1]

Walter C. Williams
University of Washington School of Medicine

The evaluation of a parents-child relationship involves the unique interaction of three distinct personalities: father, mother, and child. The child's behavior is commonly viewed as the interaction of his personality with his environment, which is determined to a great extent by the personalities and behavior of his two parents. For practical purposes, his behavior can be regarded as his reaction to his parent's behavior as he sees and interprets it.

When the child's reaction becomes disturbing to his parents, the school, or society, the family is often referred to a child guidance clinic for assistance. Here trained personnel work as a team, trying to evaluate the child's personality and behavior through interviews, "play sessions," and tests, and also trying to evaluate parental characteristics through interviews with the mother and father. Often only the mother is seen, and it is she who supplies the developmental history, presents the complaint, gives her estimate of the situation, and attempts to describe her own and her husband's personality traits.

There are several major problems involved in this traditional method. First, what are the important personal characteristics in the role of being a parent that require evaluation? Second, how can one avoid weighing the contribution of the mother so highly as to ignore the involvement of the father? Finally, how can one bridge the gap between the interviewer's evaluation of the individual parents and the perception and reactions of the child?

The more prominent dimensions upon which parental behavior can be measured are usually some variation of the following: Autocracy–Democracy, Warmth–Coldness, and Possessiveness–Detachment, suggested by the Fels group (1, 9); Authoritarian, Laissez-faire, and Democracy, suggested by Levin et al. (12); Dominant, Possessive, and Ignoring, suggested by Shoben (14). Freud (5) emphasized the oedipal pattern of castrating father with seductive mother, while Gorer (6) and Wolfenstein (17) typify the American pattern as weak father with authoritarian mother. Attempts have also been made to relate specific childhood syndromes to parental characteristics. Studies of schizophrenic children often point out the frequent occurrence of a dominating, overprotective or rejecting, immature mother, combined with a cold and intellectualizing father. Delinquent or "acting-out" children are often found to come from homes broken by death or divorce, where parents who are hostile or indifferent exercise discipline which may be classified as lax, rigid, or erratic.

The second problem, i.e., relying too heavily on the mother's interview, stems from the practical aspects of having clinic hours during the father's normal working day, making it easier to see the mother only. Much has been said of the "mother-absent" issue, both in the case of institutionalized children and the working mother, and there is increasing interest in the obvious absence of the father from the home in our current culture. A great many studies indicate that the absence of the father "makes a difference" and it seems justified to assume that his presence makes a difference also. Many clinic cases involve a general domestic discord, and the ego-defensive statements of the mother about her own, and her husband's, involvement in the child's problem cannot be taken at face value. Because of the recognized unreliability of such "self-analysis," several attempts have been made to standardize, in part, the mother's interview (13, 14), and many clinics insist that the father be seen also.

---

[1] This study was a portion of a doctoral dissertation completed at the University of Washington, 1957, supported in part through research funds of the Department of Psychiatry, University of Washington School of Medicine. Thanks are extended to Charles Strother for his valuable assistance. Thanks are also due to the staffs of Seattle Public School Guidance Department, the Archdiocese of Seattle Catholic School System, the Luther Burbank and Briscoe Memorial Schools for Boys, the Psychiatric Clinic for Children, University of Washington, and psychiatrists Benjamin Taylor and Allan Leider, all of whom assisted in the securing of subjects.

Finally, there is the difficult professional problem, after sorting out the relevant parental characteristics and situational "facts," of relating this material to the child's behavioral reactions. Between the environment, as seen by the professional interviewer, and the behavior, as reported and verified by observation in the clinic, stands a perceiving organism: the child. The immature, at times even neurotic or psychotic, child may place a quite different interpretation on his parents' behavior than would an adult, and it is to his interpretation that he is reacting. Parents have been found to differ in attitude toward children in such variables as ordinal position, sex, age levels, to mention just a few. Each parent-child relationship appears to be unique and has predictive value only in terms of the child's perception of his parents as adult sources of need-fulfillment.

The present study attempts to meet these three problems by allowing the child to evaluate his parents as he sees and reacts to them. It selects for dimensions two parental aspects recognized in our culture as important in the role of the parent: a person who "should or must be obeyed for some reason," here termed Authority; a person who is "a source of warmth and emotional support," or Love. Both mother and father are evaluated by the child in these two dimensions, high or low in Authority, high or low in Love, on a small "battery" of two tests: one a projective, one a rating scale. The scoring system is objective, having a consistent frame of theoretical reference which permits the two test forms to be compared and contrasted.

The research involves two major steps. First, a technique was developed that was suitable for use with children, measured the hypothesized parental combinations, and had satisfactory interrater reliability. Secondly, an exploratory study of this new technique was made which established that it is a practical method of distinguishing between two groups of child behavior patterns.

Dimensions Selected for Measurement

The Authority and Love dimensions selected as the basic parental characteristics to be measured were suggested by the general personality theory of Leary, Coffey, et al. (4, 7, 8, 10, 11). These researchers use a circular continuum of personality variables, using these two dimensions as horizontal and vertical axes, forming a quadrant system of High Authority–High Love (HA–HL), Low Authority–High Love (LA–HL), Low Authority–Low Love (LA–LL), and High Authority–Low Love (HA–LL). Since each dimension can be combined with the high or low of the other, they are seen as relatively independent.

In this study, the terms were simply defined as Authority meaning "one who should or must be obeyed for some reason" and Love meaning "one who is a source of warmth and emotional support." Test items were selected which would combine the high and low of each term, judged by a panel of 20 experts in the field of parent-child relations. Selection, by the child, of a preponderance of statements that the parent was one who was regarded as a person who should or must be obeyed, placed that parent in a High Authority category, whereas a preponderance of statements prejudged as low in this quality placed the parent in a Low Authority category. In like manner, the child's selection of High Love or Low Love statements placed his parent in these respective categories. Since every test item, or statement, contains these prejudged qualities of high or low in each of the two dimensions, the algebraic sum of the test items places the parent into one of the four categories: HA–HL; LA–HL; LA–LL; HA–

LL. Where algebraic summation on each dimension showed cancellation to a near zero point on both, a fifth category, the Psychologically Unknown Parent, was recorded. Since each child has two parents, each with five possibilities, the scoring system permits 25 possible parental combinations for each child.

Each of these five categories provides meaningful definitions for many of the parental characteristics suggested by the literature. Authority, exercised with Low Love (rejection) resembles "authoritarianism" or "autocracy," enforcing obedience through fear and punishment. Authority with High Love provides safeguards and guidance through respect and identification, or "democracy." Low Authority encompasses "laissez-faire" in its two aspects: with Low Love it is "ignoring"; with High Love it is "overindulgent" or "permissive." Excessively High Love and High Authority may amount to "overprotective" or "overpossessive." The Love axis tends to be the leavening agent for the Authority axis, providing a two-dimensional system that resembles Baldwin's three-dimensional analysis of parent personality (1) which also uses a central axis of "Warmth–Coldness," called "High Love–Low Love" in this study.

Development of the Measuring Instruments

Several goals were set in the initial planning of the new tests. First, a small "battery" was felt preferable, using the two types of tests found most useful in assessing interpersonal relations: the rating scale and the projective. The projective was provided with prejudged, forced selection answers, so that it could be objectively scored within the same theoretical framework as the rating scale. The cartoon form of this test provided the second goal: it made the material intrinsically interesting to children. A third goal, that all words be geared to a third grade reading level to permit self-administration, was accomplished with the assistance of current second and third grade readers and Durrell's reading lists (2). The final goal, acceptance by the three disciplines dealing with parent-child relations in a clinic setting, was accomplished by using seven psychiatrists, seven clinical psychologists, and six psychiatric social workers as the judges of the test items.[2] These 20 experts had test items submitted to them individually, were given the definitions of Authority and Love stated above, and were asked to rate each item as high or low on each dimension separately. The method of sequential experimentation was used with the first six raters, three from each discipline, called the Criterion group. These six had to agree to a criterion level of 91% to 100%, i.e., 11 of the 12 ratings, for an item to be accepted and incorporated into the test. Items not meeting this level were eliminated and new items resubmitted to this panel individually until all items met this standard. These final items were then submitted, without change, to the additional 14 raters. Final total agreement with the present scoring system, including the Criterion group's ratings, was 94% with the Rating scale form and 98% with the Projective form.

Each form of the test has 32 items for each parent, or 64 items in all. Of the 32 items in each scale, 8 fall into each of the four prerated categories of the system: HA–HL; LA–HL; LA–LL, and HA–LL. When each axis is scored separately, there are 16 possible HA statements, 16 possible LA, 16 possible HL, and 16 possible LL statements. The highest score in any of these four "characteristics" would be 16, and the lowest score would be a selection of eight high, eight low, cancelling algebraically to zero. This algebraic sum for each axis, Authority and Love, either positive or negative, can be plotted on a circular grid to give a visual presentation of the parent's position within one of the four quadrants, or close to the axis intersection. This last is, of course, the Psychologically Unknown Parent. This visual grid system is similar to the one used in the Leary-Coffey Interpersonal system (4) and resembles the standard four-quadrant grids showing ordinate and abscissa for the plotting of algebraic coordinates.

---

[2] Special thanks are due these extremely busy professional people who sacrificed valuable personal time to rate, and often rerate, items many times: psychiatrists Francis Bobbitt, Lida Brown, Ardis Candy, Richard Jarvis, Herbert Ripley, Glenn Strand, Benjamin Taylor; psychologists Sidney Bijou, Barbara Etzel, Ralph Herschstein, Otis Ramsey, Ira Steisel, Charles Strother, Theodore Tjossem; and social workers Arthur Abrahamson, Virginia Cowling, Clarice Hoekstra, Margaret Mykut, Carol Northrop, and Nancy Post.

The two forms, in combination, are called the PALS tests, short for Parental Authority–Love Statements.[3] The projective, in 16-page booklet form, devotes eight pages to a child interacting with his mother, and eight duplicate pages showing the child interacting with his father. Using blank-faced line drawings, similar to Rosenzweig's Picture-Frustration Test, the child is shown in a "Need" situation, that is, presenting a problem upon which the parent must take some action. The eight Needs shown with each parent follow Murray's formulation of basic physiological and psychological needs: Food, Sleep, Elimination, Overt Affection, Independence, Aggression, Socialization, and Succorance. These Needs are pictured as problem areas in the parent-child relationship, showing the typical "complaints" seen in clinics. Thus the Need Food shows a child refusing to eat certain foods, the Need Elimination shows a child admitting enuresis, the Need Aggression shows sibling rivalry. Four "typical" answers for the parent to make are shown below the cartoon, and represent the four major categories. These answers represent actual solutions presented by parents in a clinic setting when discussing the actual situations complained about. The child being tested reads instructions presenting these situations as "a boy and his mother and father," and is asked to circle the answer he thinks the mother or father would give. A Girl's Booklet shows a girl in these situations, and a Boy's Booklet shows a boy, aiding identification. The projective form is called PEN PALS (Projected Essential Needs, Parental Authority–Love Statements).

The rating scale is identical for both boys and girls, and is called Child's PALS. One side of the page contains 32 items of a general personality nature appropriate to fathers and has space for the child to mark each item as "Like my Father" or "Not like Father." The opposite side has 32 items appropriate to mothers and a similar arrangement to mark these statements as Like or Not like Mother. The choice is again forced, and the items are prejudged into one of the four major categories. Note, however, that the child is here told that he is rating his own parent, and his evaluative statements are much more conscious than when marking the projective. For this reason the projective was always given first in the exploratory study.

Exploratory Study

Subjects and Administration

To determine if the PALS tests could distinguish between two populations, thus indicating practical diagnostic value, two well-defined samples were selected. The known difference was a matter of public record in some social agency: one group had been classed by society as "acting-out" in a hostile and aggressive manner, whereas the other group had not been referred for this behavior but were classed by their teachers as "normal" and well-behaved, channeling their aggressions into socially acceptable competition. A maximum control was established on all other variables: the two groups were matched as to age, sex, race, type of school attended, socioeconomic status of family, and presence of two natural parents in the home. Boys, aged 9½ to 13½, white, of normal intelligence, living at home with both natural parents, were also matched as to type of occupation of fathers. Fifty children were used in each population, or one hundred in all. The hypotheses dealt with predictions that the two groups would use the PALS tests to classify their parents in a different manner: the Acting-out group as "socially undesirable" people, the Normal group as "socially desirable" people.

The Acting-out children were tested first, then matched from the wider selection of Normal children, 14% from private schools, 86% from public schools, in accordance with the proportions in the Seattle school systems. "Acting-out" was defined as having social agency or private practice referral for such socially undesirable behavior as overt aggression, stealing, lying, destructiveness, running away, or other overt revolt against adult authority. Twenty-three, mostly from the lower-income group, had been committed by court action to special schools, 23 were patients at the University's Psychiatric Clinic for Children, and represented lower to upper middle class. Four were upper class, patients of psychiatrists in private practice. Normals were matched to this group using Type and Level of parental occupation, a modification of the Index of Status Characteristics published by Warner, Meeker, and Eels (15). The mean value indicated that the matched populations clustered around the skilled and semiskilled manual occupations. The median age for both groups was 11.54, with half above and half below this value. All were in the upper four elementary school years. The Normals were individually selected with the cooperation of teachers to match the acting-out group except for the tested variable.

---

[3] Inquiries concerning copies of the tests may be addressed to the author: Walter C. Williams, University of Washington School of Medicine, Seattle 5, Washington.

All tests were administered by the experimenter who was male, white, age 42. Tests were administered individually in a private room. Standard instructions for self-administration were given, and reading ability verified by having the child read the instructions aloud. The projective, PEN PALS, was given first, followed by the rating scale, Child's PALS. Time ranged from 20 to 30 minutes for both tests, including checking to see that all items had been marked.

Scoring and Evaluation of Test Results

The scoring system, representing an algebraic summation on two axes, places the individual parent into one of the five categories. Since the variable being studied in the two groups of children dealt with the social desirability of their behavior, a similar division was logical in relating the category of the parent to the behavior of the child. The two categories indicating High Love were selected as SD, or Socially Desirable, whereas the three categories indicating rejection were classed as SU, or Socially Undesirable. The Authority axis was relegated to a secondary position, since High Authority with Love has favorable social connotations of providing the child with understanding controls, but Low Authority with Love, or Permissiveness, is also favored in our culture. Taking the parents as a "team," both should be high in Love for an SD combination. Thus, of the 25 possible parental combinations, only 4 are SD, while 21 are SU because one or both parents are classed as rejecting.

Results

Twenty-three of the 25 possible parental combinations appeared in the 200 protocols of the combined groups. The two missing patterns showed the mother low on both axes, with the father High Authority–Low Love, and Low Authority–High Love. Two combinations accounted for 60% of all ratings. Combination I, showing both parents high in both axes, accounted for 37% of all ratings. Combination II showed the father high on both axes, with the mother high in Love, low in Authority, and represented 23% of the ratings. All other patterns were below 5% of the total.

Comparing the two populations, however, presented quite a different picture. Combination I accounted for 55% of all Normals' ratings but only 19% of the Acting-out group's ratings, and Combination II was selected by 27% of the Normals and 18% of the Acting-out children. In summary, these two, showing father high in both Authority and Love, and mother high in Love but not as strong in Authority, accounted for 82% of all the Normal children's ratings, but only 37% of the Acting-out children's ratings. The remaining 63% of this last group's ratings tended to cluster around those designating the father as "Authoritarian," i.e., high in Authority, low in Love, accounting for another 48%, in contrast to only 14% of the Normals' ratings of father in this category. The Acting-out children also tended to class their mothers as Psychologically Unknown, that is, either inconsistent or inadequate in both Authority and Love. Thus 38% of their ratings of mother fell in this category, as compared to only 4% of the Normals' ratings.

Grouping the parental combination ratings for each test form, for each population, in terms of the previously established definition of Social Desirability (high in Love for both parents), Table 1 presents the observed data. Evaluating this data against an "expected" or statistically derived cell frequency of 8 SD, 42 SU, the chi square values were highly significant (beyond the .001 level) for both Normal and Acting-out children using the rating

Table 1

Number of Children in the Normal and Acting-Out Populations Rating Their Parental Combinations as Socially Desirable or Socially Undesirable, Contrasting the Use of the Rating Scale, Child's PALS, to the Projective Form, PEN PALS

	Child's PALS			PEN PALS		
	SD	SU	N	SD	SU	N
Normals	49	1	50	38	12	50
Acting-out	36	14	50	11	39	50
Totals	85	15	100	49	51	100

sheet, Child's PALS, and for the Normals using the projective PEN PALS. The data for the Acting-out children, using the projective form, was not significantly different from a chance distribution.

It was now possible to compare the populations with each other, analyzing their performances on each test form, evaluating the major hypothesis that each form had the ability to distinguish the two populations. On the rating scale, Child's PALS, the comparison of the two groups, yielded a chi square value of 11.29, significant beyond the .001 level at 1 *df*. On the projective form PEN PALS, the chi square value was even higher, 27.05, beyond the .001 level at 1 *df*. The hypothesis was supported, therefore, that on each form of the test the Normals rated their parents as socially desirable combinations more frequently than did the Acting-out children.

It was also hypothesized that the Normal children would be more consistent in their evaluation of their individual parent, whereas the Acting-out children would tend to change more often, that is, rate a parent one way on the projective, another way on the rating scale. To test this the two forms were taken as a unit: either they were both used to rate the parents as SD, or both SU, or SD on one form, SU on the other. Table 2 shows the data recorded in this fashion, a three-way division for the father and another three-way division for the mother. The term "Change" is used here to mean "absolute change," since the direction, that is, which test was SU and which SD, is ignored.

Each of the three columns, or "possibilities," can be evaluated by comparing it to the sum of the other two. Using Fisher's Exact Method (3), a probability of less than .001 was found for the difference between the two groups rating their parents as SD. The hypothesis is therefore supported that Normal children will tend to rate both mother and father as Socially Desirable consistently on both forms of the test more often than will the Acting-out children.

Using the same method, that is, combining the SD and Change to compare with the SU column, it was found that the hypothesis was supported that the Acting-out children tended to rate their parents consistently SU more often than did the Normals. Fisher's method gave a less significant value for this type of "consistency": .003 for the difference in rating fathers and .013 for the difference in rat-

Table 2

Number of Children Rating Individual Parents as Socially Desirable on Both Forms of the PALS Tests, as Socially Undesirable on Both Forms, or as Socially Desirable on One Form, Socially Undesirable on the Other

	Fathers				Mothers			
	Both SD	Both SU	Change SD–SU	*N*	Both SD	Both SU	Change SD–SU	*N*
Normals	39	0	11	50	46	0	4	50
Acting-out	14	8	28	50	21	6	23	50
Totals	53	8	39	100	67	6	27	100

ing mothers. Both are accepted as statistically significant at these levels, however.

Combining the two "Both" columns, we compare "Consistency" vs. "Change." As was predicted, the Normals were more consistent, whereas the Acting-out group more often tended to change the parent from SD on one test, then shift to SU on the other. Chi square for both the mother and father data was significant at the .001 level. All the changes shown for the father data were in the direction of SD on the Child's PALS to SU on the PEN PALS, therefore no additional analysis was necessary. The "Change" shown for the mother data involved a reversal in four of the ratings: three for the Acting-out and one for the Normals. Additional tests were run to insure that these reversals did not affect the statistical significance. They were deducted from the "Change" value, and the difference was still found to be significant at the .001 level. The reversals were evaluated by Fisher's method, and a probability of .25 was found, i.e., not significant. The only significant difference appears to be the more usual pattern of Change: rating the parent SD on the rating scale and SU on the projective.

As with the other data, chi square tests were run routinely on the two "Consistency" measures and the "Change" measure to see if they were statistically different from a chance distribution. Expectancy for Both SD is 8, for Both SU is 18, for Change is 24. All four populations were evaluated: fathers and mothers for each group. The tests for all four sets of data showed the SD ratings more frequent than chance expectations at better than the .05 level, the SU ratings less frequent than chance at the .01 level. The data for both parents, Normal group, for Change was significantly less than chance at the .01 level. The "Change" value for the Acting-out group for both parents was not significantly different from a chance distribution.

Discussion

The Testing Instruments

Although it is difficult to evaluate the clinical usefulness of a testing instrument until it has been used with many populations over an extended period of time, there are several practical aspects which favor the type of technique used in this study. Administration of both forms takes from 15 to 30 minutes, and scoring takes less than 10 minutes. The standardized scoring system, based on expert opinion, not only eliminated to a great extent the test-maker's personal bias in item selection but will continue to function in this way to eliminate the administrators' bias in test interpretation. Item selection through sequential experimentation with representatives of psychiatry, clinical psychology, and social work tended to screen out interdisciplinary differences and should increase communication and acceptance of the test results.

Of the two forms, the projective (PEN PALS) had greater discriminatory power, was less subject to social-desirability factors, and was obviously more interesting to both raters and children. Although there is a temptation to use this form without the Child's PALS, several factors indicate that the saving in time would not compensate for the loss of information. The more direct rating scale most closely resembles the type of answers given by a child in a series of psychiatric interviews when asked to describe his parents. Unlike the forced choice of the PEN PALS, the child is free to discard an item as "not like" his parent, and those he selects carry extra weight because of this factor. By itself the Child's PALS shows discriminatory power, and appears to tap the "more conscious" recognition of parental characteristics, in contrast to the "less conscious" areas expressed in the projective form. Both forms, used as a small "battery," give a more complete picture of the child's overt and covert feeling about his parents as people as well as parents, and within a consistent frame of reference for purposes of comparison. Finally, a "Consistency" or "Change" evaluation, indicated as a significant factor by analysis, depends on the use of both forms.

The Exploratory Study

The use of 23 of the 25 parental combinations possible in this scoring system would indicate that classification systems limiting parental combinations to a few cultural stereotypes may miss important relationships. The two favored combinations, on inspection, appear to be the most culturally acceptable, however, and present a measure of "conformance." The fact that 82% of the ratings by Normals, in contrast to only 37% of the Acting-out group's ratings, conformed to the cultural ideals suggests that the PALS tests are sampling an important aspect of the difficulties experienced by Acting-out children. The present study makes no claims as to where the "nonconformance" is located. The present level of interpretation places it merely in the behavior of the child on this test, but this does not preclude the strong possibiliy that additional investigation will move the level of interpretation up to the behavior of the parent. It may well be that this Acting-out child is using this test as just another means of acting-out against his parent, making it a sensitive behavioral sample. It may also be that the parent is nonconforming, at various levels of unconscious to conscious or covert to overt behavior, and that this is perceived and reacted to by the child in imitation and identification. Additional studies are now in progress to determine what levels of interpretation seem justified by the data.

Several points require clarification. The socioeconomic status represented the less controlled data of the Acting-out group and became a selective factor in the Normals, who were used as matched controls. It cannot be interpreted to mean that "in general, Normals and Acting-out children come from the same socioeconomic group." The requirement that all children come from unbroken homes was also a selective factor for both groups. It became apparent, however, that it was much more selective for the Acting-out children, since it was so much more difficult to find children of this group who did not come from broken homes. The Acting-out children selected, therefore, were probably "more normal" than an unselected group would be. A third element requires consideration: the order of test presentation. The projective form was always given first, and it is unknown how this affected the difference found in the two forms.

All hypotheses were confirmed, regarding the difference in performance of the two populations studied, at a high level of statistical confidence. Both test forms, taken as individual tests and in combination to form a "battery," were found to distinguish the populations. It is obvious that the principal contribution of this study was the development of the technique, not a thorough analysis of the differences between normal and acting-out children, nor a complete analysis of the various uses of the test forms. The limited exploratory study was used to show discriminatory power and support plans for more intensive research with additional populations. Such plans include studies of girls, of different age groups, cross-validation with other techniques using the child-centered viewpoint, having parents evaluate themselves and spouse, having expert interviewers fill out the test for the three individuals involved.

Summary

A new technique, combining a rating scale and projective with consistent theoretical framework and objective scoring system, has been developed as a research and applied clinical tool in the area of parent-child relationships. Geared to third-grade reading level, it permits a child to evaluate his unique interpersonal relationship with each parent as he sees it. He rates each parent on two continua: Authority (one who should or must be obeyed for some reason) and Love (a source of warmth and emotional support). The "high" and "low" of each used as axes form a quadrant scoring system of four major categories, plus a fifth, the "Psychologically Unknown" parent, defined in terms of algebraic cancellation.

The rating scale, Child's Parental Authority–Love Statements (Child's PALS) and the projective, Projected Essential Needs, Parental Authority–Love Statements (PEN PALS) can be visually plotted to compare and contrast the conscious and less conscious evaluation of each parent. Test items were derived by the method of sequential experimentation, using 20 expert raters from the fields of psychiatry, clinical psychology, and psychiatric social work. Interdisciplinary agreement was 94% on the Child's PALS and 98% on the PEN PALS.

A limited exploratory study was made to determine if the PALS tests could distinguish between two populations of known characteristics in a given area. The area was a record of delinquent behavior, i.e., referral to some legal or social agency for antisocial or socially unacceptable behavior. Boys with such a record were termed Acting-out, boys without such record but who were normally competitive, were termed Normal. Both groups, 50 in each, had been living with both natural parents and were matched for age, race, socioeconomic status, and type of school attended. Self-administered individual tests were given, the PEN PALS first.

Results were evaluated at a child-behavior level: samples of a known tendency of one group to act-out in a socially undesirable manner and the other group to conform in a socially desirable way. Parental combinations were thus divided into two major groups: those showing both parents high in Love, hence Socially Desirable; those showing one or both parents low in Love (rejecting) or psychologically unknown, hence Socially Undesirable.

All predictions were supported at a high level of statistical significance. Twenty-three of the possible 25 parental combinations were used, with Normals showing a much higher percentage of the two most acceptable parental patterns. These showed the father as high in both Authority and Love, with one pattern showing the mother also high in both, the other showing the mother high in Love only. Both forms, individually and in combination, showed Normals more likely to rate their parental combinations and their individual parents as Socially Desirable and Acting-out children more likely to rate their parents as Socially Undesirable. The Acting-out group also tended to change an individual parent's classification from undesirable on one test form (PEN PALS) to desirable on the other form (Child's PALS) more often than did the Normals.

It is concluded that since the PALS tests have been established as having the ability to discriminate between these two populations of boys they may also be useful in the research and diagnosis of other behavior syndromes and other populations.

Received June 27, 1958.
Early Publication.

References

1. Baldwin, A. L. *Behavior and development in childhood.* New York: Dryden, 1955.
2. Durrell, D. D. Primary remedial reading vocabulary. Mimeographed word list for lower grades and remedial reading work. Boston Univer. Educ. Clinic, 1940.
3. Fisher, R. A. *Statistical methods for research workers.* (10th ed.) Edinburgh: Oliver & Boyd, 1946.
4. Freedman, M., Leary, R., Ossorio, A., & Coffey, H. The interpersonal dimension of personality. *J. Pers.*, 1951, **20**, 143–161.
5. Freud, S. Three contributions to the theory of sexuality. In A. A. Brill (Ed.), *The basic writings of Sigmund Freud.* New York: Modern Library, 1938. Pp. 553–604.
6. Gorer, G. *The American people (A study in national character).* New York: Norton, 1948.
7. Laforge, R., Freedman, M., Leary, T., Naboisek, H., & Coffey, H. The interpersonal dimension of personality: II. An objective study of repression. *J. Pers.*, 1954, **23**, 129–153.
8. Laforge, R., & Suczek, R. The interpersonal dimension of personality: III. An interpersonal checklist. *J. Pers.*, 1955, **24**, 94–112.
9. Lasko, J. R. Parent-child relationships: Report from the Fels Research Institute. *Amer. J. Orthopsychiat.*, 1952, **22**, 300–304.
10. Leary, T. Theory and measurement methodology in interpersonal communication. *Psychiatry*, 1955, **18**, 147–161.
11. Leary, T., & Coffey, H. Interpersonal diagnosis: Some problems of methodology and validation. *J. abnorm. soc. Psychol.*, 1955, **50**, 110–125.
12. Lippitt, R., & White, R. K. Patterns of aggressive behavior in experimentally created social climates. *J. soc. Psychol.*, 1939, **10**, 271–299.
13. Sears, R. R., Maccoby, Eleanor E., & Levin, H. *Patterns of child rearing.* Evanston, Ill.: Row, Peterson, 1957.
14. Shoben, E. J. The assessment of parental attitudes in relation to child development. *Genet. Psychol. Monogr.*, 1939, **39**, 133–148.
15. Warner, W. L., Meeker, Marchia, & Eels, K. *Social class in America.* Chicago: Science Research Associates, 1949.
16. Williams, W. C. *Child's PALS and PEN PALS.* Seattle, Wash.: Author, 1957.
17. Wolfenstein, Martha. The image of the child in contemporary films. In Margaret Mead & Martha Wolfenstein (Eds.), *Childhood in contemporary cultures.* Chicago: Univer. Chicago Press, 1955. Pp. 277–293.

Journal of Consulting Psychology
Vol. 22, No. 6, 1958

Cultural Symbolism: The Age Variable

R. G. Stennett and Merle Thurlow[1, 2]
Ontario Hospital, St. Thomas, Ontario, Canada

Levy (1) reported that his experiment with 62 normal school children did not support the Freudian hypothesis of sexual symbolism. Starer (2) subsequently reported a similar experiment with male and female psychotic *S*s and student nurses, which gave strong support to this same hypothesis. The present study attempts to discover the source of the discrepancy in the results of these two investigations.

The first part of this study consisted of a replication of Starer's (2) experiment with a modification in the number of *S*s, viz., 20 psychotic adults, 10 male and 10 female, and 25 university students, 15 male and 10 female. As in Starer's original study, *S*s were asked to match 10 names (five male, five female) with 10 figures (five "male" symbols and five "female" symbols).

The number of correct matches for the psychotic and university populations were significant (chi squares 35.47 and 52.19 respectively, $p < .001$).

These results confirm Starer's findings. They do not point directly to the cause of the Starer-Levy discrepancy, but do provide additional support for the hypothesis of cultural symbolism.

In the search for the factors which might have produced the discrepancy between the Levy and Starer results, several possible variables suggested themselves: (*a*) Starer used individual testing, whereas Levy used a group-testing technique; (*b*) each author used stimulus figures of his own design; and (*c*) the subject populations differed with respect to composition (normal vs. psychotic and normal) and age (children vs. adults). However, since Starer's results with student nurses were positive and since the results of our replication of Starer's work with university students were positive, the normal vs. psychotic difference did not seem important. We also reasoned that the group vs. individual testing difference was relatively unimportant. It seemed, therefore, that the discrepant results were due to either differences in the stimulus figures or differences in the age of the populations studied. The purpose of the second part of the study was to investigate the figure variable.

*S*s in this part (37 student nurses, with a mean age of 19) were group tested with both the Levy and Starer figures, with instructions to match the names to the figures.

The *S*s matched the Levy and Starer figures with male and female names better than chance. The chi squares for the Levy and Starer figures were 13.72 ($p < .01$) and 38.90 ($p < .001$) respectively.

Since a group-testing procedure was used with both Levy and Starer figures and since *S*s matched names to both sets in a statistically significant manner, differences in the stimulus figures cannot account for the discrepancy under investigation. The variable of age, therefore, assumes crucial importance. This suggests that a careful developmental study of cultural symbolism would reveal data of theoretical importance.

Brief Report.
Received June 13, 1958.

References

1. Levy, L. H. Sexual symbolism: A validity study. *J. consult. Psychol.*, 1954, **18**, 43–46.
2. Starer, E. Cultural symbolism: A validity study. *J. consult. Psychol.*, 1955, **19**, 453–454.

---

[1] The authors wish to thank E. Starer and L. H. Levy who kindly provided copies of the figures and names used in their respective studies.

[2] An extended report of this study may be obtained without charge from R. G. Stennett, Chief Psychologist, Ontario Hospital, St. Thomas, Ontario, Canada, or for a fee from the American Documentation Institute. Order No. 5741, remitting $1.25 for microfilm or $1.25 for photocopies.

Journal of Consulting Psychology
Vol. 22, No. 6, 1958

PSYCHOLOGICAL TEST REVIEWS

New Test

Cattell, Raymond B., Beloff, Halla, & Coan, Richard W. *IPAT High School Personality Questionnaire (H.S.P.Q.)*. Ages 12–17 years. 2 forms. Untimed, (40) min. Test booklet Form A or Form B ($4.00 per 25 or $3.50 per 2 or more packages); answer sheet ($1.90 per 50 or $14.90 per 500); hand scoring key (60¢); handbook, pp. 58 ($2.20), and tabular supplement of norms, pp. 10 (80¢); sample set $3.10). Champaign, Ill.: Institute for Personality and Ability Testing, 1958.

The *H.S.P.Q.* is a thorough revision of the earlier *IPAT Junior Personality Quiz,* and is a close relative of the *Sixteen Personality Factor Questionnaire.* Each form of the *H.S.P.Q.* contains 140 forced-choice items, 10 items for each of 14 factors. No item is scored for more than one factor. The 14 factors are drawn from the well-known factor-analytic studies of the senior author, and include Schizothymia vs. Cyclothymia, Intelligence, Neuroticism vs. Ego-strength, Phlegmatic Temperament vs. Excitability, etc. The provision of two forms is an unusual asset among personality questionnaires. The handbook recommends that one form be used for survey testing with the second form for retesting, or that both forms be used for more intensive individual studies so as to increase reliability.

Data in the handbook bring to sharp focus an issue emphasized by the test's authors—broad coverage with few items per factor versus more reliable coverage of a smaller number of variables. Judged by usual psychometric standards, the reliabilities of the individual factor scores are low. Retest with one form after two weeks yields factor reliabilities of from .52 to .66 with a median of .58. The correlations of the factor scores obtained from the two forms are even lower, from .26 to .51, with a median of .37. The latter data certainly raise questions about the equivalence of the forms and about the success with which the factors have been represented by the items. On the other hand, the construct validities, in terms of the multiple correlations of the items with the pure factors as obtained from the original factor analyses, are reported as fairly high, ranging from .58 to .78 for each form, and from .73 to .88 for the two forms used together. Intercorrelations are generally low, with only 5 of 91 *r*'s exceeding .30.

Tentative norms in stens (10-point standard scores) and deciles are based on 1,189 American boys and girls, ranging in age from 11 to 20, with a median of 14.5 years. Their geographic distribution is unspecified, but they seem to come mainly from a few midwestern cities. Tables are given for boys and for girls separately, and for the sexes combined. Age corrections are supplied for the two factor scores which vary significantly with age.

The authors urge that no factor score be interpreted separately. Two methods for global use are described, by matching profiles and by regression equations. Unfortunately, these useful methods are represented only by illustrative examples. There is no compendium of criterion profiles or of regression equations against external criteria. The two illustrations of the regression method—for the prediction of leadership and of school achievement—are so inadequately described that the reader is unable to evaluate the merit of the research on which they are based. In contrast to the implied rigor of the regression equation method, the handbook contains several pages of "clinical" suggestions, supported by no explicit data.

In summary, the reviewer sees the *H.S.P.Q.* as a potentially valuable instrument when further research has supplied sound actuarial data. But it is subject to serious misuse if the unreliable single factor scores are given stereotyped interpretations, supported only by the interpreter's intuitive hunches about the meanings of the factors.—*L. F. S.*

Erratum

The review of the *Minnesota Counseling Inventory* in the June issue of this Journal (p. 241) contained the misstatement that "there is no indication of item analysis against a criterion." Although no detailed data are given concerning the item analyses, the Manual (p. 24) states that all items were evaluated against the external criteria of nominations and ratings, and against the internal criterion of total score on the appropriate scale. The Editor is grateful to Dr. Wilbur L. Layton for calling his attention to this oversight.—*L. F. S.*

Brief Reports

The *Journal of Consulting Psychology* will accept Brief Reports of research studies in clinical psychology for early publication without expense to the author. The procedure is intended to permit the publication of soundly designed studies of specialized interest or limited importance which cannot now be accepted because of lack of space. Several pages in each issue will be devoted to Brief Reports, published in the order of their receipt without respect to the dates of receipt of the regular articles. Most Brief Reports appear in the first or second issue to go to press following their final acceptance.

An author who wishes to submit a Brief Report:

1. Sends the Brief Report, limited to one printed page and prepared according to the specifications given below.

2. Also sends to the Editor a full report of the research study, in sufficient detail to give a clear account of its background, procedure, results, and conclusions, which will be filed with the American Documentation Institute to insure indefinite availability.

3. Prepares at least 100 mimeographed copies of the full report, which the author will send without charge to all who request it as long as the supply lasts.

4. Agrees not to submit the full report to another journal of general circulation.

Specifications

Brief Report. The Brief Report should give a clear, condensed summary of the procedure of the study and as full an account of the results as space permits.

To insure that the Brief Report will be no longer than one printed page, its typescript, including all matter except the title and the author's lines, must not exceed 75 lines averaging 42 characters and spaces in length. Set the typewriter margins for short lines of 42 characters, which are 3.5 inches long in elite typing, and 4.2 inches long in pica.

The manuscript of the Brief Report must be *double spaced* throughout. Except for its short lines, it follows the standard style (1). Headings, tables, and references are avoided or, if essential, must be counted in the 75 lines. Each Brief Report must be accompanied by a footnote in the style below, which is typed on a separate sheet and *not* counted in the 75-line quota: [1]

[1] An extended report of this study may be obtained without charge from John Doe, 300 Market St., Prospect 6, Mass. (giving the author's full name and address), or for a fee from the American Documentation Institute. Order Document No. ——, remitting $—— for microfilm or $—— for photocopies.

Extended report. Because the extended report is intended for photoduplication, and is not copy to be sent to a printer, its style should differ in several ways from that of other manuscripts: (*a*) The extended report should be typed with single spacing for economy in duplication. (*b*) Tables and figures should be placed adjacent to the text which refers to them. A caption should be typed below each figure. (*c*) Footnotes should be typed at the bottom of the page on which reference is made to them. In other respects, the full report is prepared in the style specified by the *Publication Manual* (1).

Reference

1. American Psychological Association. Council of Editors. *Publication manual of the American Psychological Association (1957 rev.).* Washington, D. C.: American Psychological Association, 1957.